DRAFT: Category Theory for Computer Science

J.R.B. Cockett [1]
Department of Computer Science, University of Calgary,
Calgary, T2N 1N4, Alberta, Canada

October 12, 2016

[1]Partially supported by NSERC, Canada.

Contents

Chapter 1

Basic Category Theory

This chapter contains the basic elements of category theory.

1.1 The definition of a category

A **category** is a mathematical object and as such has a precise definition: one can therefore establish whether something is or is not a category. There are, however, various ways to define a category which are "morally" equivalent. We shall start with the most common presentation of a category which is as a directed graph with composition. Another important presentation of a category is as an enriched category: we shall concentrate on Set-enriched categories, although it is this view which allows us to understand the structure of Cat the category of categories.

1.1.1 Categories as graphs with composition

A **category**, $\mathbb{C}$ consists of a directed graph, that is a collection of **objects**, $\mathbb{C}_0$, and a collection of **maps**, C_1, such that each map $f \in \mathbb{C}_1$ has an associated object $D_0(f) \in \mathbb{C}_0$ which is its **domain** and an associated object $D_1(f) \in \mathbb{C}_0$ which is its **codomain**. We shall indicate that f has domain A and codomain B by writing $f : A \longrightarrow B$.

The maps of a category can be composed: that is given any pair of maps $f, g \in \mathbb{C}_1$ with the codomain of f being the same as the domain of g, that is they are a **composable** pair of the form $f : A \longrightarrow B$ and $g : B \longrightarrow C$, then there is an associated **composite** $fg : A \longrightarrow C$.[1] The operation of forming composites is called **composition**. Categories are required to have identity maps for the composition. Thus, given any object $C \in \mathbb{C}$ there is associated to it an **identity** map, denoted by $1_C : C \longrightarrow C$. This data must then satisfy the following axioms:

[C.1] (Identity laws) if $f : A \longrightarrow B$ then $1_A f = f = f 1_B$,

[C.2] (Associative law) if $f : A \longrightarrow B$, $g : B \longrightarrow C$, and $h : C \longrightarrow D$.

Now when checking that something is actually a category one must first establish that all the data is present. Thus, one must first identify a collection of objects. It will often be the case that

[1] Please note that "diagrammatic" as opposed to "applicative" order will be used in this text. Applicative order is particularly common in many of the more mathematical text books on category theory so you should be able to read both.

the objects will be known by some other name such as nodes, types, sets, or algebras. Next one must have a collection of maps. Again the maps will often be known by some other name such as arrows, terms, functions, or homomorphisms. Finally one must have a notion of composition for these maps: this involves not only having the operation which allows the formation of composites for maps which juxtapose but also the presence of identities.

We remark that having identities is a property rather than extra structure as an identity (if it exists at an object) is uniquely determined by the composition. This is because if 1_A and $1'_A$ are identities on A then $1_A = 1_A 1'_A = 1'_A$; so there can be at most one identity map.

The identity map is an example of a map which starts and ends at the same place: these we call **endo-maps**. A category can have – and often will have – many endo-maps.

A **subcategory** $\mathbb{C}'$ of a category $\mathbb{C}$ is given by any subcollection of the objects and maps which is "closed" to the domain, codomain, identity, and composition structure.

Please notice that it has *not* been required that the objects or arrows of a category should form a *set*: this is because some categories are too large for either their objects or their arrows to form sets. A typical "big" category is the category of sets Set itself: Russell's paradox[2] informs us that the set of all sets cannot possibly be a set. These issues have given rise to much philosophical and mathematical work which, here we will not discuss. Suffice it to say that one way of retrieving the situation is to allow for a hierarchy of set theories (a class system) each one containing the previous set theory and the "class" of all sets of the previous set theory. This allows one to view the category of sets to be a category in the *next* set theory up in the hierarchy.

1.1.2 Categories as partial semigroups

Another way to view a category is as collection of maps which have a partial associative multiplication and a system of units. This view is interesting as it indicates that the objects are actually redundant structure and their role can be replaced by the identity maps. However, the cost of making them redundant is that one needs some additional axioms.

This time a category $\mathbb{C}$ just consists of a collection of maps on which there is a partial associative composition and two assignments of arrows to arrows D_0 and D_1 such that:

[C'.1] $D_0(f)f$ is always defined and is f,

[C'.2] $fD_1(f)$ is always defined and is f,

[C'.3] fg defined if and only if $D_1(f) = D_0(g)$ and $D_0(fg) = D_0(f)$ and $D_1(fg) = D_1(g)$,

[C'.4] (Associative) $(fg)h = f(hg)$ whenever each side is defined,

[C'.5] $D_j(D_i(x)) = D_i(x)$, for $i, j = 0, 1$.

To return to the definition above it is necessary to provide the collection of objects: these are the maps for which $D_0(x) = x = D_1(x)$ which are also the identity maps.

[2] Russell's paradox concerns the set of all sets: this should be a set because as one can define sets by properties (such as being a set) but consider the set of all sets which are not members of themselves - a subset of the set of all sets. If it were a member of itself then it could not be a member of itself; on the other hand if it were not a member of itself then it would be a member of itself! This paradox forced set theorists to declare that the set of all sets cannot be a set. As one can hardly deny the existence of this collection of all sets, the awkward question then arises: what is the set of all sets?

1.1.3 Categories as enriched categories

It is useful to introduce the following notation, which also leads naturally into our third view of what a category is: if $\mathbb{C}$ is a category then by $\mathbb{C}(A, B)$ will be denoted the collection of all arrows with domain A and codomain B. These are called the **hom-objects** of the category. If these collections are actually *sets* then we call these "hom-sets" and say that the category is **enriched** in Set.

Notice that when a category is enriched in Set, the composition can be described as a family of maps, one for each triple of objects in $\mathbb{C}$:

$$m_{A,B,C} : \mathbb{C}(A, B) \times \mathbb{C}(B, C) \longrightarrow \mathbb{C}(A, C)$$

(where m is for multiplication). In addition, there must be an identity map $1_A : 1 \longrightarrow \mathbb{C}(A, A)$, where 1 as an object represents a generic one element set $1 = \{*\}$. These maps must now satisfy the two requirements imposed on a category (associativity and the identity laws):

$$\begin{tikzcd}
\mathbb{C}(A,B) \times \mathbb{C}(B,C) \times \mathbb{C}(C,D) \arrow[r, "m_{ABC} \times 1"] \arrow[d, "1 \times m_{BCD}"'] & \mathbb{C}(A,C) \times \mathbb{C}(C,D) \arrow[d, "m_{ACD}"] \\
\mathbb{C}(A,B) \times \mathbb{C}(B,D) \arrow[r, "m_{ABD}"'] & \mathbb{C}(A,D)
\end{tikzcd}$$

$$\begin{tikzcd}
1 \times C(A,B) \arrow[r, "1_A \times 1"] \arrow[dr, "\pi_1"'] & \mathbb{C}(A,A) \times \mathbb{C}(A,B) \arrow[d, "m_{AAB}"] \\
& \mathbb{C}(A,B)
\end{tikzcd}
\qquad
\begin{tikzcd}
C(A,B) \times 1 \arrow[r, "1 \times 1_A"] \arrow[dr, "\pi_0"'] & \mathbb{C}(A,B) \times \mathbb{C}(B,B) \arrow[d, "m_{ABB}"] \\
& \mathbb{C}(A,B)
\end{tikzcd}.$$

These requirements I have expressed as commuting diagrams - a style which we shall see much more of as we progress.

This enriched category style of expressing the structure of a category is important as often the hom-objects of categories are not simply unstructured collections of maps but themselves have structure (e.g. Abelian categories, poset enriched categories, Cat-enriched categories, etc.). By regarding a category in this manner we will be able to explain how the additional structure of the hom-objects must interact with the composition of the category itself.

For enriched categories a very natural notion of subcategory is determined by taking a subcollection of the objects and leaving the hom-objects and composition structure the same. This is called a **full-subcategory**.

1.1.4 The opposite category and duality

Category theory is full of symmetries which are called **dualities**. The basic source of symmetry is the ability to reverse arrows. Thus, given any category we may obtain a new category by keeping everything the same except to switch the direction of the arrows. If we start with a category $\mathbb{C}$ and flip the direction of the arrows we obtain a new category $\mathbb{C}^{\mathrm{op}}$. Observe now that anything which is true of $\mathbb{C}$ now holds in the **dual** form in $\mathbb{C}^{\mathrm{op}}$. Thus, when we prove a result there is always another result, obtained by reversing the sense, of the arrows which will also be true. This principle of duality allows us to get double the bang for our buck!

Often the prefix "co" (as in colimit, coequalizer, coproduct) is an indication that this is the dual concept to that with this prefix removed.

Some categories are actually **self-dual**, thus $\mathbb{C}^{\mathrm{op}}$ is in some sense the same as $\mathbb{C}$. In this case there is often an explicit translation (actually this is an example of a contravariant functor see section 1.5.1) $(\_)^* : \mathbb{C} \longrightarrow \mathbb{C}^{\mathrm{op}}$ this has the following properties:

(i) If $f : A \longrightarrow B$ then $f^* : B^* \longrightarrow A^*$;

(ii) If $f : A \longrightarrow B$ and $g : B \longrightarrow C$ then $(fg)^* = g^* f^*$ and $1_A^* = 1_{A^*}$.

We shall say that such a translation is an **involution** in case $(f^*)^* = f$ for each arrow. Sometimes it is the case that for objects $A = A^*$ in this case we shall refer to the involution as a **converse**.[3]

1.1.5 Examples of categories

Below we outline a series of sources of examples of categories.

Preorders

Categories enriched in sets of cardinality at most 1 are called **preorders**. They are important as they account for both equivalence relations and partially ordered sets. There is a considerable body of knowledge about partially ordered sets and this provides an important sources of examples and, often, a pattern to follow when developing categorical structures.

Notice when there is at most one arrow between any two objects the value of the composite of any two maps is forced. Thus, it is simply a matter of whether maps exists between objects or not. Thus we may view such a category as a relation on the objects.

A preorder may also be viewed as a category enriched in Set_1, the category of sets with at most one element (in other words the category containing, as objects, the empty set and the one element set, and, as maps, all set maps between these objects).

A **preordered set** is set with a reflexive, transitive relation . A relation is **reflexive** on a set X in case whenever $x \in X$ we require (x, x) to be in the relation: one may think of this categorically as giving the identity map on that object. A relation is **transitive** in case whenever (x, y) and (y, z) are in the relation then (x, z) must be in the relation: one may view this categorically as giving the composition.

A relation is an **equivalence relation** in case in addition it is **symmetric** that is whenever (x, y) is in the relation (y, x) is also in the relation. This is equivalent to asking that every arrow is an isomorphism (see section 1.2.1: a category in which all the maps are isomorphisms is called a groupoid).

Clearly an equivalence relation viewed as a category has a converse involution.

A **partially order set** is a preorder with the addition **anti-symmetry** property that whenever (x, y) and (y, x) are in the relation then $x = y$. This is equivalent categorically to asking either that the only isomorphisms are the identity maps or, less stringently, that all isomorphisms are endo-maps. This is a small illustration of how one notion for a poset can be generalized in different ways for arbitrary categories.

[3]Notice that an involution cannot in general be stationary on maps. If an involution is stationary on maps one can easily show that there is only one object in each connected component of the category and that the endomorphism monoid on each object is a commutative monoid.

Finite categories

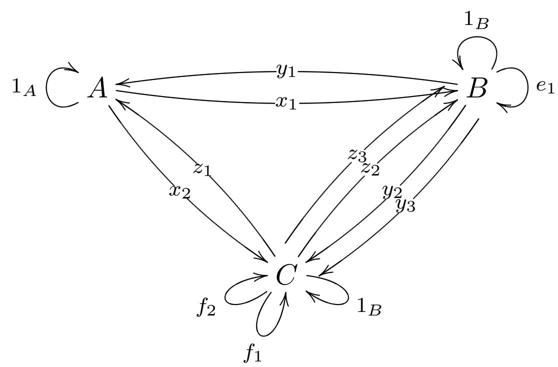
The simplest category of all has no objects and no maps. This is called (for reasons which will be explained later) the **initial** category. The initial category is certainly finite and there is not much else one can say about it!

The next most simple category is the category with one object and exactly one arrow. This is called the **final category**: it is also finite and there is not so very much more one can say about it either. The one arrow is actually forced to be the identity map on the one object.

A finite category must have both a finite number of objects and a finite number of arrows. All finite categories are necessarily enriched over finite sets (i.e. have finite hom-sets).

Before we leave finite categories let us look at how we may present them. A finite category as a directed graph together with a multiplication tables for each object: each table will represent the arrows which juxtapose at that object.

We may view a finite category $\mathbb{F}$ as having an underlying directed graph such as the one below:



We may then arrange the composition tables by the objects at which the arrows meet. For each such table we may arrange the arrows coming in to the object (along the vertical axis) according to the object from which they come. Similarly, along the horizontal axis we may arrange the maps according to the object at which they end. The result of the composition (minimally) must then be an arrow with the correct domain and codomain. One then must check that the composition satisfies the identity and associative laws. The latter law is quite arduous to check as one must consider triples of composable maps and check that the two possible compositions are in fact equal.

Finite categories are a great source of examples and counter-examples for simple categorical facts. Notice also that there are two simple ways of constructing new finite categories from old ones. Given two categories $\mathbb{C}$ and $\mathbb{D}$ one can form the disjoint union of their arrows and objects to form a new category $\mathbb{C} + \mathbb{D}$, the coproduct, where the composition(s) are unchanged. Alternately one can put the compositions in parallel to form $\mathbb{C} \times \mathbb{D}$ the **product** category. Here one takes the cartesian product of the objects and of the arrows and define the composition pointwise, that is if $f : C_1 \longrightarrow C_2$ in $\mathbb{C}$ and $g : D_1 \longrightarrow D_2$ in $\mathbb{D}$ then in $\mathbb{C} \times \mathbb{D}$ there is the maps $(f, g) : (C_1, D_1) \longrightarrow (C_2, D_2)$ and $(f, g)(f', g') = (ff', gg')$.

Categories enriched in finite sets

A category all of whose homsets are finite is a category enriched in finite sets. Every finite category is a finite set enriched category. However, as the category of finite sets, Set_f, is certainly finite set enriched not every fiite set enrched category is a finite category.

		A	B	C
	A	1_A	x_1	x_2
A	1_A	1_A	x_1	x_2
B	y_1	y_1	e_1	y_3
C	z_1	z_1	z_3	f_2

		A	B		C	
	B	y_1	1_B	e_1	y_2	y_3
A	x_1	1_A	x_1	x_1	x_2	x_2
B	1_B	y_1	1_B	e_1	y_2	y_3
	e_1	y_1	e_1	e_1	y_3	y_3
C	z_2	z_1	z_2	z_3	f_1	f_2
	z_3	z_1	z_3	z_3	f_2	f_2

		A	B		C		
	C	z_1	z_2	z_3	1_C	f_1	f_2
A	x_2	1_A	x_1	x_1	x_2	x_2	x_2
B	y_2	y_1	1_B	e_1	y_2	y_2	y_3
	y_3	y_1	e_1	e_1	y_3	y_3	y_3
C	1_C	z_1	z_2	z_3	1_C	f_1	f_2
	f_1	z_1	z_2	z_3	f_1	f_1	f_2
	f_2	z_1	z_3	z_3	f_2	f_2	f_2

Figure 1.1: Composition tables for $\mathbb{F}$

Another example of a finite set enriched category is the category of finite interference graphs, Intf_f. This category has object finite sets with a symmetric, anti-reflexive relation $(S, \bowtie)$. Explicitly the relation satisfies:

- $x \bowtie y \Leftrightarrow y \bowtie x$,
- $x \bowtie y \Rightarrow x \neq y$.

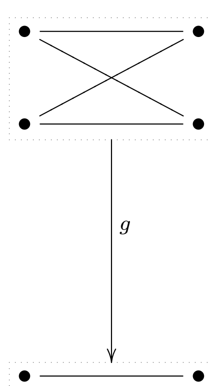
A map between two objects in this category

$$f : (X, \bowtie_X) \longrightarrow (Y, \bowtie_Y)$$

is a map between the underlying finite sets such that

$$x \bowtie x' \Rightarrow f(x) \bowtie f(x').$$

Here is an example of a map which collapses the four cycle C_4 onto the completely separated graph S_2:



As we shall discover this category, and its brother Intf which allows infinite interference graphs, has alot of interesting structure.

Sets

There are various categories we may form from the category of sets. The primary category, denoted by Set, is the category of sets and functions. However, there are two other variants which we shall want to use as examples: the category of relations, Rel, and the category of partial maps, Par.

The category of relations Rel is given by the following data:

Objects: Sets;

Maps: Relations $R : X \longrightarrow Y$;

Composition: $RS = \{(x,z)|\exists y.(x,y) \in R \wedge (y,z) \in S\}$;

Identities: $1_X = \{(x,x)|x \in X\}$.

The category of sets and partial maps, Par, and the category of sets and functions, Set, itself, can be seen as subcategories of this category. The category of partial maps is the subcategory with the same objects but only those relations which are **definite**, that is (x,y) and (x,y') imply $y = y'$. The category of functions is a further subcategory of the category of partial maps Par with the same objects but with the maps restricted to those definite relations which are total. A relation $R : X \longrightarrow Y$ is **total** in case for every $x \in X$ there is a $y \in Y$ such that $(x,y) \in R$.

Notice that neither of these subcategories are full-subcategories as we are strictly reducing the set of maps between objects (in most cases!).

The category of relations has a converse $(\_)^o : \mathsf{Rel}^{\mathrm{op}} \longrightarrow \mathsf{Rel}$ obtained by reversing the ordered pairs. This converse operation, however, is not inherited by either the category Par or Set. The relations in Rel may also be ordered by inclusion: this makes the category poset enriched. This enrichment *is* inherited by Par: for Set this partial order is discrete (in the sense that two things are related only if they are the same).

Monoids

A category which has one object and a set of maps is a **monoid** in Set. Monoids have been extensively studied in their own right. For example group theory is the study of monoids all of whose maps are isomorphisms (see section 1.2.1) and these besides being monoids are examples of categories.

There are large numbers of finite groups, larger numbers of finite monoids and, thus, by implication there are even larger numbers of finite categories!

Vector spaces

Let R be any **rig** (this a ring without the "n" for negatives), that is a set with an addition (a commutative associate operation $x + y$ with an identity 0) and a multiplication (that is an associative not necessarily commutative operation $x \cdot y$ with a unit 1) such that

$$x \cdot (y + z) = x \cdot y + x \cdot z \quad \text{and} \quad (y + z) \cdot x = y \cdot x + z \cdot x,$$

$$x \cdot (y \cdot z) = (x \cdot y) \cdot z \quad \text{and} \quad x \cdot 1 = x = 1 \cdot x$$

$$x \cdot 0 = 0 = 0 \cdot y$$

then we may form $\mathsf{Mat}(R)$ the category of R–matrices. This category has objects the natural numbers and maps $n \times m$–matrices with the usual multiplication.

Somewhat unusually we also allow $0 \times n$ and $n \times 0$ matrices. The composites with these "empty" matrices are themselves empty.

Notice that $\mathsf{Mat}(R)$ has a converse involution given by transposition.

A special example of this category which is very well-studied is $\mathsf{Mat}(K)$ where K is a field (such as $\mathbb{R}$).

Path categories

Given a directed graph $\mathcal{G}$ we may form a category from it which we call the **path** category of $\mathcal{G}$, denoted $\mathsf{Path}(\mathcal{G})$. The objects are the same as those of $\mathcal{G}$ but the arrows are sequences of arrows in $\mathcal{G}$ which juxtapose. The composition is, as might be expected, given by concatenation.

More formally, the arrows in $\mathsf{Path}(\mathcal{G})$ are given by triples $(A, [g_0, .., g_n], B)$ where, when the list of maps is non-empty, $A = D_0(g_0)$, $B = D_1(g_n)$, and $D_1(g_i) = D_0(g_{i+1})$ for $i = 0, ..., n-1$. When the list is empty we insist that $A = B$ which gives us maps of the form $(A, [], A)$ for each object A: these will serve as the identity maps. We define $D_0(A, l, B) = A$ and $D_1(A, l, B) = B$ and set $(A, l_1, B)(B, l_2, C) = (A, l_1@l_2, C)$, where $\_@\_$ is the usual concatenation operation.

$\mathsf{Path}(\mathcal{G})$ is an important category as we shall learnt it is the *free* category on a graph. However, there are also some important modification of this example which are of traditional importance in Computer Science: as a preview of things to come, we briefly outline these modifications.

The first modification is to allow the arrows from A to B to be (regular) subsets of the set of all paths from A to B. This means that besides concatenation we need to allow the formation of unions of sets of paths and of the Kleene star of all endo-path sets. The composition of one set of paths with another is given by taking the composite of all possible pairs.

In this category $f(g \cup h) = fg \cup fh$ and $(f \cup g)h = fh \cup gh$. This allows us to do a matrix construction and to form $\mathsf{Reg}(\mathcal{G})$:

Objects: Lists of objects of $\mathcal{G}$, e.g. $[G_1, .., G_n]$

Maps: Matrices of regular subsets

$$[a_{i,j}] : [G_1, .., G_n] \longrightarrow [G'_1, .., G'_m]$$

where $a_{i,j} : G_i \longrightarrow G'_j$.

Composition: Matrix multiplication where multiplication is composition and addition is union:

$$[a_{i,j}][b_{j,k}] = [\bigcup_j a_{i,j} b_{j,k}]$$

Identities: Identity matrices with the singleton set consisting of the identity path on the diagonal and empty sets off diagonal.

Those familiar with the translation from a finite automaton into a regular language will know there is an interesting additional operation (expresed using the Kleene star) on these matrices which allows one to reduce the dimension across a common input and output state: esentially this operation allows "feedback" and makes that state "internal" to the matrix.

For a 2-dimensional matrix this operation works as follows:

$$\mathrm{trace}_X \left(\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} : [A_1, A_2, X] \to [B_1, B_2, X] \right)$$

$$= \begin{bmatrix} a_{11} \cup a_{13}a_{33}^*a_{31} & a_{12} \cup a_{13}a_{33}^*a_{32} \\ a_{21} \cup a_{23}a_{33}^*a_{31} & a_{22} \cup a_{23}a_{33}^*a_{32} \end{bmatrix} : [A_1, A_2] \to [B_1, B_2]$$

where $(\_)^*$ is the Kleene star operation. This operation is the basis of the translation of finite state machines into regular languages. The idea is to view a finite state machine (with ϵ-transitions) as a matrix in this category: the domain is arranged to consist of a distinguished start state and the internal states while the codomain is a distinguished final state and the internal states. The operation of removing states can then be used to eliminate rows and columns until one obtains a regular expression from start state to final state.

We shall see the Kleene star is equivalent to having the categorical structure of a *trace* (see Chapter **??**).

Programming languages

We may think of a programming language as being a given by a collection of types which have programs defined between them. The ability to compose programs and the presence of the "do nothing" program together with the expectation that these will satisfy the basic axioms of a category means that programming languages may be modeled by categories.

Given a programming language there then is the question of which category precisely it describes. Given that many programming languages have grown in an ad hoc way this is likely, in general, to be a messy question. However, one might expect that for simple programming languages this would be an easy question to answer. Now there is always the "term model" to fall back on, that is the programming language itself with the equalities that should hold. However, this is regarded, rightly, as a less than satisfactory answer: instead an answer which allows the language to be modeled in some other well-understood mathematical structure is sought.

These are are generally known as **semantic** issues: formally one is seeking a functor with certain properties from the programming language into a category (the "semantics") which is independently constructed and understood. One usually would require the functor to be full or faithful: sometimes as program equality is not well-understood one takes equality to be "semantic equality" and, thus, the main concern becomes the fullness.

Surprisingly this task turns out to be far from simple. In fact, it has only been with the advent of game theoretic models that solutions to these semantic issues have even begun to emerge. These semantics based in game theory are actually very subtle combinatorial models and they reflect programming constructs in a surprisingly faithful manner. Furthermore, it is clear that game theoretic models provide a rich source of material for providing constructions of free categories with various constructs.

More sceptically, these game theoretic models may also be viewed as simply a combinatorial re-expression of syntax. Thus, it might be argued, that the, perhaps Platonically inspired, task of providing a completely independent semantics has not really been fulfilled. To a mathematician – and particularly a category theorist, these philosophical issues, of course, have an extremely hollow ring. Mathematics is after all about finding connections between concepts: necessarily things which are close together are connected!

The pursuit of semantics is a major theoretical direction in programming languages. However, there is another direction which is considerably less traditional but, nonetheless, of greater practical importance. One can ask which constructs should be present in a good programming language. This is a seemingly rather vague question, however, one can give a reasonable answer to such a question if one understands the theory of programming constructs. Category theory provides a powerful mathematical tool for addressing this sort of question: it gives a landscape view of the constructs involved in programming and how they can be fit together to make reasonable programming environments.

1.1.6 Exercises

(1) Prove that the first two definitions of a category are equivalent. Explain why the third is not quite equivalent (i.e. what precisely is the effect of being enriched over Set).

(2) Provide an example of a preorder which has an involution which is not a converse.

(3) Check that the finite category $\mathbb{F}$ is really a category. Is $\mathbb{F}^{\mathrm{op}}$ the "same" category as $\mathbb{F}$ (if indeed $\mathbb{F}$ is a category!).

(4) The category $\mathbf{2}$ is

$$\begin{tikzcd} A \arrow[loop left, "1_A"] \arrow[rr, "a"] & & B \arrow[loop right, "1_B"] \end{tikzcd}$$

What do the categories $\mathbf{2} + \mathbf{2}$ and $\mathbf{2} \times \mathbf{2}$ look like?

(5) Write a program to find all (non-isomorphic) categories with n-objects and m-arrows. How many categories are there with 1,2,3,4, and 5 arrows?

(6) Show that $\mathsf{Path}(\mathbb{C})$ is a category.

(7) Prove carefully that $\mathsf{Reg}(\mathcal{G})$ is a category.

(8) Provide an example of a category enriched over finite sets which is *not* a finite category.

(9) Here is an illustration of how two categories can have the same objects and maps but a completely different composition structure. Consider sets with relations but alter the composition to be:

$$RS = \{(x, z) | \forall y.(x, y) \in R \vee (y, z) \in S\}.$$

Prove that this forms a category (hint: what are the identities).

(10) Do the total relations form a subcategory of Rel?

(11) Prove that $\mathsf{Mat}(\mathbb{R})$ is a category as defined and that transposition is really a converse involution. If you are ambitious show that $\mathsf{Mat}(R)$ is a category in which transposition is a converse whenever R is a rig.

(12) (Harder) Show that $\mathsf{Set}^{\mathrm{op}}$ is the category of atomic Boolean algebras with (arbitrary) meet and join preserving maps.

1.2 Basic properties of maps

A map in a category can have a number of properties. The most basic of these are outlined in this section. In this section we also begin to see value of the dualities which are present in category theory.

1.2.1 Epics, monics, retractions, and sections

A map $f : A \longrightarrow B$ in a category $\mathbb{C}$ is **monic** (short for a *monomorphic* map – sometimes abbreviated to "mono") in case whenever $k_1 f = k_2 f$ then $k_1 = k_2$. Dual to the notion of a monic map is that of an epic map: a map $f : A \longrightarrow B$ in a category $\mathbb{C}$ is **epic** (short for epimorphic map – sometimes abbreviated to epi) in case whenever $fh_1 = fh_2$ then $h_1 = h_2$. The fact that a map is monic does not stop it from being epic as well: a map that is both epic and monic we shall refer to as being **bijic**.

A map $f : A \longrightarrow B$ is a **section** in case there is a map $f' : B \longrightarrow A$ such that $ff' = 1_A$. Dual to a section is a retraction: a map $g : A \longrightarrow B$ is a **retraction** in case there is a map $g' : B \longrightarrow A$ such that $g'g = 1_B$. It is quite possible for a map to be both a section and a retraction: such a map is called an **isomorphism**. Clearly identity maps are always isomorphisms.

The following gives some basic facts concerning these properties:

Lemma 1.2.1 *In any category $\mathbb{C}$ it is the case that:*

(i) *The composition of monics is monic;*

(ii) *The composition of epics is epic;*

(iii) *If fg is monic then f is monic;*

(iv) *If fg is epic then g is epic;*

(v) *All sections are monic;*

(vi) *All retractions are epic;*

(vii) *The compositions of sections is a section;*

(viii) *The composition of retractions is a retraction;*

(ix) *If fg is a section then f is a section;*

(x) *If fg is a retraction then g is a retraction.*

Isomorphisms are rather special: if $f : A \longrightarrow B$ is a map we shall refer to a map $g : B \longrightarrow A$ such that $fg = 1_A$ as the **right inverse** of f. Similarly a map $h : B \longrightarrow A$ such that $hf = 1_B$ will be referred to as the **left inverse** of f. A section has a right inverse and is itself a left inverse, while a retraction is a right inverse and has a left inverse. An isomorphism has both a left inverse and a right inverse:

Lemma 1.2.2 *If $f : A \longrightarrow B$ has a left inverse h and a right inverse g then $h = g$.*

PROOF: Observe f is both epic and monic as it is both a section and a retraction. Thus, $fh = fh1_A = fhfg = f1_Bg = fg = 1_A$ so that h is also a right inverse of f. But then $fh = fg$ and as f is epic $h = g$. □

As the inverse of an isomorphism f is unique we shall denote it f^{-1}. We have the following alternative characterizations of isomorphisms:

Lemma 1.2.3 *The following are equivalent:*

(i) *f is an isomorphism;*

(ii) *f is an epic section;*

(iii) *f is a monic retraction.*

PROOF: If f is an isomorphism then the remaining characterizations are immediate. As these characterizations are dual it suffices to complete the proof to show that an epic section is an isomorphism. It suffices to show that f is a retraction: let f' be a right inverse of f then $ff'f = f$ now, as f is epic, this implies $f'f = 1_B$ showing f is a retraction. □

1.2.2 Idempotents

An endo-map $e : A \longrightarrow A$ is an **idempotent** if $ee = e$. Notice that if $h : A \longrightarrow B$ is a retraction with left inverse $h' : B \longrightarrow A$ then hh' is an idempotent as $hh'hh' = h1_Bh' = hh'$. We shall say that an idempotent e is **split** if there is a retraction h with left inverse h' such that $e = hh'$. While an idempotent may be split in many different ways there is an important a sense in which there is essentially just one splitting:

Lemma 1.2.4 *Suppose $e : A \longrightarrow A$ is an idempotent and $h_1 : A \longrightarrow B_1$ has left inverse h'_1 and $h_2 : A \longrightarrow B_2$ has left inverse h'_2, with $e = h_1h'_1 = h_2h'_2$ then there is a unique isomorphism $k : B_1 \longrightarrow B_2$ such that $h_1k = h_2$ and $kh'_2 = h'_1$.*

PROOF: Set $k = h'_1h_2$ then the required identities hold. Furthermore k is an isomorphism as $k^{-1} = h'_2h_1$ (as $h'_1h_2h'_2h_1 = h'_1eh_1 = h'_1h_1h'_1h_1 = 1_{B_1}$ and similarly for the other composite $h'_2h_1h'_1h_2 = 1_{B_2}$). Suppose k' also satisfies $h_1k' = h_2$ then $h_1k' = h_1k$ and as h_1 is a retraction and therefore epic it follows that $k = k'$. □

We shall say that the splitting of an idempotent is *unique up to unique isomorphism.* It is certainly not the case that idempotents will generally split in a category, however, there is an important construction which allows one to freely split idempotents.

Let $\mathbb{C}$ be any category. Let $\mathsf{Split}(\mathbb{C})$ be the following category:

Objects: Idempotents e of $\mathbb{C}$;

Maps: $(e_1, f, e_2) : e_1 \longrightarrow e_2$ where $e_1 : A \longrightarrow A$ and $e_2 : B \longrightarrow B$ is a map $f : A \longrightarrow B$ in $\mathbb{C}$ such that $e_1fe_2 = f$;

Compositions As in $\mathbb{C}$ on the middle coordinate: $(e_1, f, e_2)(e_2, g; e_3) = (e_1, fg, e_3)$.

Identities: the identity for an idempotent is that idempotent $(e,e,e) : e \longrightarrow e$.

Now it is not hard to show that this is a category. What is interesting about this category is that all the idempotents in it split:

Proposition 1.2.5 *Let $\mathbb{C}$ be any category then $\mathbb{C}$ is a full subcategory of* $\mathsf{Split}(\mathbb{C})$ *and all idempotents split in* $\mathsf{Split}(\mathbb{C})$.

PROOF: We may regard $\mathbb{C}$ as a full subcategory of Split by letting the identity maps (which are certainly idempotent) represent the objects of $\mathbb{C}$ in $\mathsf{Split}(\mathbb{C})$.

Suppose $(e,k,e) : e \longrightarrow e$ is an idempotent in $\mathsf{Split}(\mathbb{C})$ then k is an idempotent in $\mathbb{C}$. But then we have maps $(e,k,k) : e \longrightarrow k$ and $(k,k,e) : k \longrightarrow e$ in $\mathsf{Split}(\mathbb{C})$ and it is easy to check that these provide a splitting for (e,k,e). □

To motivate why this construction may be of more than passing interest consider the category of partial recursive functions on the natural numbers, Rec. Each enumerable set may be characterized by an idempotent which is the computation which returns the element unchanged when it is in the recursively enumerable set but simply does not terminate on elements outside. In $\mathsf{Split}(\mathsf{Rec})$ there is an object or type for each enumerable set. Thus, this gives an example of a unityped system from which can be constructed a very rich type system.

1.3 Finite set enriched categories and full retractions

Categories enriched in finite sets have a number of rather special properties: a notable one is that they always have a fully retracted "skeleton". An object is **fully retracted** in case its only idempotent endomorphism is the identity map. In a finitely enriched category – in which idempotents split – every object has, up to isomorphism, a unique full retract. The subcategory of fully retracted object is the *fully retracted skeleton* of the category.

Recall that a category is enriched in finite sets in case it is an ordinary (Set-enriched) category in which all the homsets are finite sets. This does not mean the number of objects is finite as the category of finite sets, Set_f, is certainly finite set enriched yet by no means has a finite number of objects. Indeed, any Set_f-concrete category (i.e. a category with a faithful functor to finite sets) will be finite set enriched so that the category of finite groups, rings or fields are all finite set enriched.

A peculiar property that finite set enriched categories have is that every endo-map if iterated to a high enough power will start to repeat itself. This allows an idempotent to be associated to each endo-map: a category in which one can associate to each endo-map an idempotent in this manner is said to be **retractive**. If, in addition, each object has an associated fully retracted object the category is said to be **fully retractive**.

1.3.1 Retractive inverses

A **retractive inverse**, $g : B \longrightarrow A$, for a map $f : A \longrightarrow B$ is a map such that $gfg = g$. To explain why this is called a "retractive" inverse consider the following situation

$$A \xleftarrow{s} S \xrightarrow{s'} B$$

where s and s' are sections and so have right inverses r and r' respectively, then rs' has $r's$ as a retractive inverse as $r'srs'r's = r's$. Thus, whenever two objects have a retract in common they will be connected by a map which has a retractive inverse. Notice in this case, although this does not happen in general, also $r's$ has a retractive inverse rs'. In this case we shall say that the pair of maps are **mutual retractive inverses**.

When idempotents split, being connected by a map with a retractive inverse implies that the domain and codomain of the map have a common retract. To see this first note that if f has a retractive inverse g then both fg and gf are idempotents and in the idempotent splitting

$$g : gf \longrightarrow fg \quad \text{and} \quad fgf : fg \longrightarrow gf$$

but also $g(fgf) = gf$ and $(fgf)g = fg$ so that these make these idempotents isomorphic. This means we have a pair of sections

$$1_A \xleftarrow{fg} fg \xrightarrow{fg} 1_B$$

exhibiting a common retract of the objects.

Notice that if g is a retractive inverse for f then f need not be a retractive inverse of g. However fgf will be a retractive inverse of g as $(fgf)g(fgf) = fgf$. Thus when g is a retractive inverse of f then g and fgf are always mutual retractive inverses.

Retractive inverses need not be unique but they do enjoy the following weak uniqueness property: if g and g' are retractive inverses of f such that $fg = fg'$ and $gf = g'f$ then $g = g'$, as $g = gfg = gfg' = g'fg' = g'$. This is as might be expected as gf and fg are, from our analysis, supposed to split via the same object (upto isomorphism).

An endomorphism $h : A \longrightarrow A$ has an **central retractive inverse** g in case g is a retractive inverse of h such that $hg = gh$: the "central" prefix refers to the fact that the two idempotents generated are the same. Notice that this means that $g^n h^n = (gh)^n = gh$ and $h^n g^n = (hg)^n = hg$ and has the consequence that if g and g' are central retractive inverses of f with $g^n = g'^n$ for some $n \geq 1$ then $g = g'$. This as:

$$\begin{aligned} g &= gfg = (gf)^n g = g^n f^n g = g'^n f^n g = (g'f)^n g = g'fg \\ &= g'(fg)^n = g'f^n g^n = g'f^n g'^n = g'(fg')^n = g'fg' = g' \end{aligned}$$

The central retractive inverses of h can be ordered: suppose g_1 and g_2 are such then $g_1 \leq g_2$ if $g_1 h g_2 = g_2 = g_2 h g_1$. This is clearly a reflexive relation. It is transitive as if $g_1 \leq g_2$ and $g_2 \leq g_3$ then

$$g_1 h g_3 = g_1 h g_2 h g_3 = g_2 h g_3 = g_3 \quad \text{and} \quad g_3 h g_1 = g_3 h g_2 h g_1 = g_3 h g_2 = g_3$$

and antisymmetric as if $g_1 \leq g_2$ and $g_2 \leq g_1$ then $g_1 = g_1 h g_2 = g_2$.

A central retractive inverse r of $h : A \longrightarrow A$ is a **least central retractive inverse** in case for any other central retractive inverse g of f we have $r \leq g$ (that is $gfr = g = rfg$). Clearly if h has a least central retractive inverse it must be unique.

Clearly any automorphism has as its least retractive inverse its ordinary inverse. Any idempotent e has its least reflexive retractive inverse e. To see this first we note that a reflexive retractive inverse of an idempotent is always an idempotent as $g = gegeg = ggeeg = ggeg = gg$. e is clearly a reflexive retractive inverse of itself but for any other such inverse g we have

$$eeg = eegeg = geeeg = geg = g = geg = geeeg = gegee = gee$$

so that e is the least central retractive inverse.

We shall say that a category is **retractive** in case each endomorphism $h : A \longrightarrow A$ has a least central retractive inverse r_h such that $r_{fg} f = f r_{gf}$.

Clearly every groupoid is a retractive category, but there is also an important source of examples in finitely enriched categories:

Proposition 1.3.1 *Any finite set enriched category is retractive.*

PROOF: Suppose $f : A \longrightarrow A$ then there are smallest numbers $k, h, h', m > 0$ such that $f^k = f^{k+h}$ and $m \cdot h = k + h'$. Set $r_f = f^{2 \cdot m \cdot h - 1}$ then certainly $f r_f = r_f f$ but also

$$\begin{aligned} r_f f r_f &= f^{2 \cdot m \cdot h - 1} f f^{2 \cdot m \cdot h - 1} \\ &= f^{2 \cdot m \cdot h} f^{k + h' - 1} \\ &= f^{k + 2 \cdot m \cdot h} f^{h' - 1} \\ &= f^{k + h' - 1} \\ &= r_f \end{aligned}$$

To show r_f is least we suppose we have a central retractive inverse g we must show that $r_f f g = g = g f r_f$. As r_f and g commute it suffices to show $r_f f g = g$ for this we have:

$$\begin{aligned} g &= gfg = (gf)^{2 \cdot m \cdot h} g \\ &= g^{2 \cdot m \cdot h} r_f f g \\ &= g^{2 \cdot m \cdot h} r_f f r_f f g \\ &= (gf)^{2 \cdot m \cdot h} r_f f g \\ &= g f r_f f g \\ &= r_f f g. \end{aligned}$$

It remains to show that $r_{fg} f = f r_{gf}$. For this we observe:

$$(fg)^{2 \cdot m \cdot h - 1} f = f (gf)^{2 \cdot m \cdot h - 1}$$

so that if $r_{gf} = (gf)^{2 \cdot m \cdot h - 1}$ we are done.

Lemma 1.3.2 *In any category if fg repeats with cycle length h after step k (i.e. we have $(fg)^k = (fg)^{k+h}$) then gf repeats with cycle length h as well and it starts repeating at or before $k+1$ steps.*

PROOF: If fg starts repeating at k then for any $k' > 1$ we have

$$(gf)^{k+k'+h} = f(fg)^{k+h+(k'-1)} g = f(fg)^{k+(k'-1)} g = (gf)^{k+k'}.$$

As this works for all $k' > 1$ it follows gf repeats no later than $k+1$. $\square$

Now $(gf)^{2 \cdot m \cdot h - 1} = (gf)^{2 \cdot (k + h') - 1}$ if the cycle length of gf is less or equal to k (that is k or $k-1$) we are done. If the cycle length is $k+1$ however and $h' = 1$ we must use the fact that

$$(gf)^{2 \cdot (k+h') - 1} = (gf)^{2 \cdot ((k+1)+h) - 1} = r_{gf}.$$

This completes the proof of proposition 1.3.1. $\square$

1.3.2 Fully retracted objects

An object is fully retracted in case its only idempotent endomorphism is the identity map.

Lemma 1.3.3 *In a retractive category:*

(i) *If two fully retracted objects are connected (that is there are maps both ways between them) then all maps between them are isomorphisms.*

(ii) *The endomorphisms of a fully retracted object form a group.*

(iii) *Any two fully retracted objects which are retracts of the same object are isomorphic.*

PROOF: Suppose A and B are fully retracted and $f : A \longrightarrow B$ and $g : B \longrightarrow A$ then $r_{fg}fg = fgr_{fg} = 1_A$ so fg is an isomorphism. This means f is a section. But similarly gf is an isomorphism so f is a retraction as well and so is an isomorphism.

Two fully retracted objects which are retractions of the same object are connected so isomorphic. □

We shall call a category **fully retractive** in case the category is retractive and every object can be fully retracted.

Lemma 1.3.4 *An object in a retractive category has a full retraction if and only if there is an idempotent e which splits such that any other idempotent e' with $ee' = e'e$ has $ee' = e$.*

PROOF: The splitting of e gives a full retracted object as any idempotent on that object would induce an idempotent e' which commutes with e on the original object and would have $ee' = e'$. □

Corollary 1.3.5 *Every finite set enriched category in which idempotents split is fully retractive.*

PROOF: The number of idempotents on an object is finite. Define a preorder on idempotents by $e \leq e'$ if $ee' = e$. This is clearly reflexive and it is transitive as $e \leq e' \leq e''$ maens $ee' = e$ and $e'e'' = e'$ so that $ee'' = (ee')e'' = e(e'e'') = ee' = e$. This preorder must have least elements: pick such a least element e_0. Now suppose $ee_0 = e_0e$ then $e_0e = e_0$ as e_0 is minimal. Thus e_0 exhibits the property required by lemma 1.3.4. □

A fully retracted object is **rigid** in case, in addition to being fully retracted, the only automorphism is the identity. In a retractive category this means, in addition, that the only endomorphism is the identity.

Proposition 1.3.6 *In a retractive category an object has a rigid full retraction if and only if there is an idempotent e which is a retractive inverse for each endomorphism of A.*

PROOF: If an object in a retractive category has a full retract which is rigid then observe that each endomorphism $f : A \longrightarrow A$ must have a retractive inverse which is the idempotent which gives the full retraction. To see this suppose $(s, r) : R \longrightarrow A$ is a full retraction of A which splits the idempotent e and R is rigid. It follows that $sfr : R \longrightarrow R$ must be the identity and thus $efe = rsfrs = rs = e$.

Conversely, suppose this property holds, namely that e is a retractive inverse for each endomorphism f then, consider $(s,r) : R \longrightarrow A$ a splitting of e and an isomorphism $\alpha : R \longrightarrow R$ then $r\alpha s$ has $er\alpha se = e$. However, then $r\alpha s = rs$ showing $\alpha = 1_R$ as r is epic and s is monic. □

A category is said to have **binary products** (see chapter 3 on limits) in case given any two objects A and B there is a third object, C, with maps $\pi_0 : C \longrightarrow A$ and $\pi_1 : C \longrightarrow B$ such that for any pair of maps $f : X \longrightarrow A$ and $g : X \longrightarrow B$ there is a unique map $\langle f, g\rangle : X \longrightarrow C$ such that $\langle f, g\rangle\pi_0 = f$ and $\langle f, g\rangle\pi_1 = g$.

In a retractive category $\mathbb{X}$ which has binary products we shall say that a full-subcategory, $\mathbb{F}$, is **down-closed** in case: whenever $X \in \mathbb{F}$ and there is a map $X \longrightarrow Y$ then $Y \in \mathbb{F}$. A down-closed full subcategory is **principal**, if if is generated by one object.

Proposition 1.3.7 *In a fully retracted category with binary products the principal down-closed full subcategories are in bijective correspondence to the isomorphism classes of fully retracted objects.*

Proof: If a down-closed full subcategory is generated by an object A then the filter must contain and be generated by any retraction of A. Therefore, one can use a fully retracted subobject to generate the filter. On the otherhand, suppose two different fully retracted objects generate a filter then they must be connected and, therefore, be isomorphic. □

The principal filters of a fully retractive category are partially ordered and this gives a partial ordering on the fully retracted objects which corresponding to the partial order given by the existence of morphisms between them.

Remark 1.3.8 The intersection of two principal filters is principal, provided the category has *coproducts*, see the chapter on limits, as it is generated by the full retraction of the coproduct of the two generating fully retracted objects. Similarly the join of the two filters is generated by the full retraction of the *product* of the two fully retracted objects.

1.3.3 Exercises

(1) Show that in Set

  (a) a map f is monic if an only if it is **injective** ($f(x) = f(y)$ implies $x = y$);
  (b) a maps f is epic if and only if it is **surjective** (for every y in the codomain there is an x such that $f(x) = y$);
  (c) all epics are retractions;
  (d) not all monics are sections (hint: consider the empty set);
  (e) all bijics are isomorphisms.

(2) Give example(s) of finite categories in which:

  (a) Not all monics are sections;
  (b) Not all epics are retractions;
  (c) Not all bijics are isomorphisms;
  (d) Not all idempotents are split.

(3) Prove lemma 1.2.1.

(4) Describe the epics, monics, sections, and retractions in Rel.

(5) Describe the epics, monics, sections, and retractions in Par.

(6) Describe the epics, monics, sections, and retractions in $\mathsf{Path}(\mathcal{G})$.

(7) (Hard!) Describe the epics, monics, sections, and retractions in $\mathsf{Reg}(\mathcal{G})$.

(8) Prove that if an idempotent is either epic or monic then it is the identity map.

(9) Prove that $\mathsf{Sub}(A)$, the category of subobjects of A, defined for an object $A \in \mathbb{C}$ as the category:

    Objects: monics $m : A' \longrightarrow A$;

    Maps: $f : m_1 \longrightarrow m_2$ maps in $\mathbb{C}$ such that $fm_2 = m_1$;

    Identities: $1_{A'} : m \longrightarrow m$ as in $\mathbb{C}$;

    Composition: As in $\mathbb{C}$.

    is a preorder.

(10) Given an example of two idempotents e_1 and e_2 such that neither e_1e_2 nor e_2e_1 are idempotents. Show that if $e_1e_2 = e_2e_1$ (the idempotents commute) then the composite e_1e_2 is an idempotent.

(11) A monoid M is commutative if for every element x and y it is the case that $xy = yx$. A monoid is a semilattice if, in addition, every element is idempotent. Describe the maps of $\mathsf{Split}(M)$ when M is a semilattice.

(12) Do all idempotents split in Rel? (Hard: give a description of the category $\mathsf{Split}(\mathsf{Rel})$ - hint: what is a completely distributive lattice! Look this up!).

(13) (Hard!) What do the idempotents in $\mathsf{Reg}(\mathcal{G})$ look like? (Open problem!) Give a description of $\mathsf{Split}(\mathsf{Reg}(\mathcal{G}))$.

(14) Find some (or all) of the fully retracted objects in:

    (a) Finite sets and maps?

    (b) Finite G-sets for a group?

    (c) Finite groups and homomorphisms?

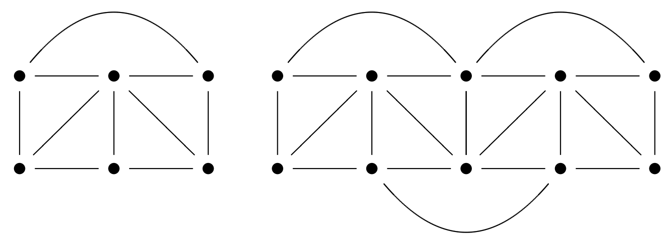
(15) The category of finite interference graphs is fully retractive by the results proven in this section. Recall that an interference graph is n-colorable if and only if it has a map to the "chaotic" interference graph on n-elements. Thus to decide n-colorability amounts to deciding the presence of such a map.

    A well-known (solved) problem is the four color theorem. Show that one can reduce the problem to showing that there are no *planar complete* fully retracted planar graphs with five nodes or more by the following steps:

    (a) Prove that an interference graph is n-colorable if and only if it's full retraction is.

(b) The four color problem is concerned only with *planar* graphs: prove that a full retraction of a planar graph is always planar.

(c) It suffices therefore to consider the four colorability of fully retracted planar graphs. If these where finite in number the four color problem would be easy! Show that this is *not* the case by considering family of interference graphs:



(d) Say that a planar graph is *planar complete* in case there is no edge which can be added without destroying the property of being planar. Are the above graphs planar complete?

(e) Prove that planar graphs are four colorable if and only if planar complete graphs are four colorable. Conclude that planar graphs are four colourable if and only if no planar complete graphs with more than five elements is fully retracted!

1.4 Orthogonality and Factorization

In the category of Set a fundamental property is the ability to factorize each map into an epic (surjective) map followed by a monic (injective) map. This, in fact, is a property which is enjoyed by many categories "built" from Set. Our objective in this section is to develop the theory of factorization systems.

In general, for a factorization system, it is not necessary for the classes of maps to have any special relationship to epics and monics. A crucial property that they must satisfy is the orthogonal condition and this is where we start with our development.

1.4.1 Orthogonal classes of maps

A map $f : A \longrightarrow B$ is **left orthogonal** to a map $g : C \longrightarrow D$ (or equivalently $g : C \longrightarrow D$ is **right orthogonal** to $f : A \longrightarrow B$) if for all maps h_1 and h_2, such that the outer square below commutes, there is a unique $k : B \longrightarrow C$ such that the two triangles below commute:

$$
\begin{tikzcd}
A \arrow[r, "f"] \arrow[d, "h_1"'] & B \arrow[d, "h_2"] \arrow[dl, dotted, "k"] \\
C \arrow[r, "g"'] & D
\end{tikzcd}
$$

Notice that left orthogonal and right orthogonal are dual notions

Remark 1.4.1 The requirement that the cross arrow k is unique is sometimes too strong. We shall say that f is **weakly left orthogonal** to g in case the cross map exists but is not necessarily unique. Many of the results we shall prove generalize to this weaker notion. Furthermore, as various notions of (process) simulation may be described using this notion it is of some importance in computer science: in algebraic topology weak factorization systems play a fundamental role in Quillen model structures.

Consider the following examples: :

(1) In Set the maps which are right orthogonal to the map

$$k : \{0,1\} \longrightarrow \{*\}; \begin{array}{ccc} 0 & \mapsto & * \\ 1 & \mapsto & * \end{array}$$

are precisely the monomorphisms (injections).

(2) In Set the maps which are *weakly* right orthogonal to the initial map

$$z_1 : \{\} \longrightarrow \{*\}$$

are precisely the epimorphisms (surjections).

Let $\mathcal{A}$ be an arbitrary collection of maps in $\mathbb{C}$ then we may form the set of maps which are right orthogonal to all the maps in $\mathcal{A}$, we shall call this collection $\mathcal{A}_\perp$. Similarly we may form the collection of maps which are left orthogonal to $\mathcal{A}$ which we shall denote ${}_\perp\mathcal{A}$:

Lemma 1.4.2 *If $\mathbb{C}$ is any category and $\mathcal{A}$ is any collection of maps then*

(i) *$\mathcal{A}_\perp$ contains all isomorphism;*

(ii) *Any $x : X \longrightarrow Y$ which is in both $\mathcal{A}$ and $\mathcal{A}_\perp$ is an isomorphism;*

(iii) *$\mathcal{A}_\perp$ is closed to composition;*

(iv) *If gh is in $\mathcal{A}_\perp$ and h is monic then g is in $\mathcal{A}_\perp$;*

(v) *If gh and h are in $\mathcal{A}_\perp$ then g is in $\mathcal{A}_\perp$.*

PROOF:

(i) An isomorphism is obviously right and left orthogonal to all maps.

(ii) The trivial commuting square

$$\begin{tikzcd} X \arrow[r, "x"] \arrow[d, equal] & Y \arrow[d, equal] \\ X \arrow[r, "x"'] & Y \end{tikzcd}$$

has a reverse diagonal which gives the inverse of x.

(iii) Suppose g_1 and g_2 are orthogonal to f and there is a commuting square

$$\begin{tikzcd} A \arrow[rr, "f"] \arrow[d, "h_1"'] & & B \arrow[d, "h_2"] \\ X \arrow[r, "g_1"'] & Y \arrow[r, "g_2"'] & Z \end{tikzcd}$$

then viewing it slightly differently we obtain a cross map k_2 as follows:

$$\begin{tikzcd} A \arrow[r, "f"] \arrow[d, "h_1 g_1"'] & B \arrow[d, "h_2"] \arrow[dl, dotted, "k_2"] \\ Y \arrow[r, "g_2"'] & Z \end{tikzcd}$$

Now as g_1 is orthogonal to f, there is a unique cross maps k_1 in

$$\begin{tikzcd} A \arrow[r, "f"] \arrow[d, "h_1"'] & B \arrow[d, "k_2"] \arrow[dl, dotted, "k_1"] \\ X \arrow[r, "g_1"'] & Y \end{tikzcd}$$

which has the desired property of a cross map for the original square.

(iv) We must show that g is orthogonal to f, so suppose $k_1 g = f k_2$ then this happens if and only if $k_1 g h = f k_2 h$. However, this square has a unique cross map v where $fv = k_1$ and $vgh = k_2 h$, now as h is monic we have $vg = k_2$.

(v) We must show that g is right orthogonal to f. As above, suppose $k_1 g = f k_2$ then as $k_1 g h = f k_2 h$ we have a unique cross map v such that $fv = k_1$ and $vgh = k_2 h$:

$$\begin{tikzcd} A \arrow[rr, "f"] \arrow[dd, "k_1"'] & & B \arrow[d, "k_2"] \arrow[ddll, dotted, "v"] \\ & & D \arrow[d, "h"] \\ C \arrow[r, "g"'] & D \arrow[r, "h"'] & E \end{tikzcd}$$

Now consider the following square:

$$\begin{tikzcd} A \arrow[r, "f"] \arrow[d, "k_1"'] & B \arrow[d, "k_2"] \arrow[ddl, dotted] \\ C \arrow[d, "g"'] & D \arrow[d, "h"] \\ D \arrow[r, "h"'] & E \end{tikzcd}$$

as $h \in \mathcal{A}_\perp$ it follows that there is a unique cross map for the square. However, both vg and k_2 will serve, thus they are equal.

$\square$

There is a dual result for the maps which are right orthogonal to a class $\mathcal{A}$ of maps whose statement we leave to the reader.

This means that if we start with some class $\mathcal{A}$ we may form $\mathcal{A}_\perp$ and we can then form ${}_\perp(\mathcal{A}_\perp)$. Notice that $\mathcal{A} \subseteq {}_\perp(\mathcal{A}_\perp)$ and one might think that one can continue in this manner to enlarge $\mathcal{A}$, by

$$\mathcal{A} \subseteq {}_\perp(\mathcal{A}_\perp) \subseteq {}_\perp(({}_\perp(\mathbb{A}_\perp))_\perp) \subseteq ....$$

however, this sequence stops after the first step once one has the largest class of maps orthogonal to all the maps orthogonal to $\mathcal{A}$.

The argument which show this is deceptively simple and uses standard facts about Galois connections. We know that $\mathcal{A} \subseteq {}_\perp(\mathcal{A}_\perp)$ for any $\mathcal{A}$ and similarly that $\mathcal{B} \subseteq ({}_\perp\mathcal{B})_\perp$: substituting $\mathcal{A}_\perp$ for $\mathcal{B}$ we obtain $\mathcal{A}_\perp \subseteq ({}_\perp(\mathcal{A}_\perp))_\perp$. However, applying $(\_)_\perp$ to $\mathcal{A} \subseteq {}_\perp(\mathcal{A}_\perp)$ reverses the inequality to give $\mathcal{A}_\perp \supseteq ({}_\perp(\mathcal{A}_\perp))_\perp$. This means $\mathcal{A}_\perp = ({}_\perp(\mathcal{A}_\perp))_\perp$ which in turn shows that the sequence becomes stationary.

Lemma 1.4.3 *Given a class $\mathcal{A}$ of maps there is a largest class of maps containing $\mathcal{A}$, namely ${}_\perp(\mathcal{A}_\perp)$, which is left orthogonal to the largest class which is right orthogonal to $\mathcal{A}$, namely $\mathcal{A}_\perp$.*

We shall call $\mathcal{A}_\perp$ and ${}_\perp(\mathcal{A}_\perp)$ the **maximal orthogonal** classes of maps generated by $\mathcal{A}$.

1.4.2 Introduction to factorization systems

We are now ready to introduce the general notion of factorization systems and establish some of their basic properties.

A **factorization system** $(\mathcal{E}, \mathcal{M})$ on a category $\mathbb{C}$ consists of two classes of maps $\mathcal{E}$ and $\mathcal{M}$ such that:

[F.1] $\mathcal{E}$ and $\mathcal{M}$ contain all the isomorphisms,

[F.2] $\mathcal{E}$ and $\mathcal{M}$ are closed to composition,

[F.3] Every map $f : A \longrightarrow B$ can be factorized into $f = e_f m_f$ where $e_f : A \longrightarrow E_f$ is in $\mathcal{E}$ and $m_f : E_f \longrightarrow B$ is in $\mathcal{M}$,

[F.4] If $f = em = e'm'$ where $e, e' \in \mathcal{E}$ and $m, m' \in \mathcal{M}$ then there is a unique isomorphism k such that $ek = e'$ and $km = m'$.

The following are some basic consequences of this definition:

Lemma 1.4.4 *If $\mathbb{X}$ is a category with an $(\mathcal{E}, \mathcal{M})$-factorization system then*

(i) *$\mathcal{E}$ is left orthogonal to $\mathcal{M}$;*

(ii) *Commutative squares $fh_2 = h_1g$ have unique factorizations that is*

$$\begin{tikzcd}
A \arrow[r, "e_f"] \arrow[d, "h_1"'] & E_f \arrow[r, "m_f"] \arrow[d, dotted, "k"] & B \arrow[d, "h_2"] \\
C \arrow[r, "e_g"'] & E_g \arrow[r, "m_g"'] & D
\end{tikzcd}$$

there is a unique map k making the smaller squares commute.

PROOF:

(i) For $\mathcal{M}$ to be orthogonal to $\mathcal{E}$ there must be a unique cross map k for any square with upper horizonal map in $\mathcal{E}$ and lower horizontal map in $\mathcal{M}$:

$$\begin{tikzcd} A \arrow[r, "e"] \arrow[d, "f"'] & B \arrow[d, "g"] \arrow[dl, dotted, "k"'] \\ C \arrow[r, "m"'] & D \end{tikzcd}$$

By factorizing the two vertical maps we obtain an isomorphism α by uniqueness of factorization:

$$\begin{tikzcd} A \arrow[r, "e"] \arrow[d, "e(f)"'] & B \arrow[d, "e(g)"] \\ E \arrow[r, dotted, no head, "\alpha"'] \arrow[d, "m(f)"'] & E' \arrow[d, "m(g)"] \\ C \arrow[r, "m"'] & D \end{tikzcd}$$

So we may set $k = e(g)\alpha^{-1}m(f)$. It remain to show this k is unique. Suppose that if k' is an alternative map then we may factorize these maps to obtain a unique isomorphism β such that $ee(k)\beta = ee(k')$ and $\beta m(k')m = m(k)m$. Also, as $km = g$ and $k'm = g$ there is also a unique isomorphism β_1 such that $e(k)\beta_1 = e(k')$ and $m(k)m = \beta_1 m(k')m$. Now observe that $\beta_1 = \beta$ as it also serves also as a comparison for the first factorization. Similarly, as $ek = f$ and $ek' = f$ there is also a unique isomorphism β_2 such that $\beta_2 m(k') = m(k)$ and $ee(k)\beta_2 = e(k')$. However, as before we may argue $\beta_2 = \beta$. This gives $k = e(k)m(k) = e(k)\beta m(k') = e(k')m(k') = k'$, as desired.

(ii) This is a straightforward application of the orthogonality.

□

There are various different ways to formulate a factorization system: an important way involves specifying just one of the classes of maps. For a given class of maps, $\mathcal{E}$, a map f has a **maximal $\mathcal{E}$-factorization** in case it can be factored as $f = ef'$ where $e \in \mathcal{E}$ and so that the following (modified orthoginality) property holds: in any commutative square

$$\begin{tikzcd} X \arrow[rr, "g"] \arrow[d, "h_1"'] & & Y \arrow[d, "h_2"] \arrow[dl, dotted] \\ A \arrow[r, "e"'] & E \arrow[r, "f'"'] & B \end{tikzcd}$$

with $g \in \mathcal{E}$ there is a unique cross map. We shall say a category has a **maximal $\mathcal{E}$-factorization** if every maps has a maximal $\mathcal{E}$-factorization.

Dually we shall speak of a category having a **maximal $\mathcal{M}$-cofactorization**.

Proposition 1.4.5 *The following are equivalent for a category $\mathbb{C}$:*

(i) *There are two classes of maps $\mathcal{E}$ and $\mathcal{M}$ such that:*

   (a) *Each $\mathcal{E}$-map is left-orthogonal to each $\mathcal{M}$-map;*

   (b) *The $\mathcal{E}$-maps are closed to composition on the right with isomorphisms (that is when α is an isomorphism and $f \in \mathcal{E}$ then, when defined, the composite $f\alpha \in \mathcal{E}$);*

   (c) *The $\mathcal{M}$-maps are closed to composition on the left with isomorphisms;*

   (d) *Each map can be factorized as $f = em$ where $e \in \mathcal{E}$ and $m \in \mathcal{M}$;*

(ii) *$(\mathcal{E}, \mathcal{M})$ are a maximally orthogonal pair of classes of maps which allows each map $f : A \longrightarrow B$ to be factorized into $f = em$ with $e \in \mathcal{E}$ and $m \in \mathcal{M}$;*

(iii) *$\mathcal{E}$ is a class of maps containing all isomorphisms and closed to composition such that each map $f : A \longrightarrow B$ has a maximal $\mathcal{E}$-factorization $f = ef'$.*

(iv) *$\mathcal{M}$ is a class of maps containing all isomorphisms and closed to composition such that each map $f : A \longrightarrow B$ has a maximal $\mathcal{M}$-cofactorization $f = f'm$.*

(v) *$(\mathcal{E}, \mathcal{M})$ is a factorization system.*

Proof: Notice first that *(iii)* and *(iv)* are dual so it suffices to prove the equivalence with just one of these: we shall choose the first.

(i)* $\Rightarrow$ *(ii) We must show that under these conditions $(\mathcal{E}, \mathcal{M})$ are maximally orthogonal. By symmetry it suffices to show that if g is left-orthogonal to $\mathcal{M}$ then $g \in \mathcal{E}$. Toward this factorize such a g as $g = em$ then we have

$$\begin{tikzcd} A \arrow[r, "g"] \arrow[d, "e"'] & B \arrow[d, equal] \arrow[dl, dotted, "k"'] \\ E \arrow[r, "m"'] & B \end{tikzcd}$$

showing that m is a retraction. But also now the following commutes:

$$\begin{tikzcd} A \arrow[r, "e"] \arrow[d, "e"'] & E \arrow[d, "m"] \arrow[dl, "mk"'] \\ E \arrow[r, "m"'] & B \end{tikzcd}$$

which, by uniqueness of cross maps, means $mk = 1_E$, showing m is an isomorphism. But now it follows $g = em \in \mathcal{E}$, as $e \in \mathcal{E}$ and m is an isomorphism, completing the proof.

(ii)* $\Rightarrow$ *(iii) Using lemma 1.4.2 we know $\mathcal{E}$ is closed to composition and contains all isomorphisms. We also know we can factorize each map: it only remains to show that this is a maximal $\mathcal{E}$-factorization. However, this follows immediately from orthogonality.

We notice, in this case, that if $f = e'f'$ where $e' \in \mathcal{E}$ then by orthogonality we have a unique k such that

$$\begin{tikzcd} A \arrow[r, "e'"] \arrow[d, "e"'] & E' \arrow[d, "f'"] \arrow[dl, dotted, "k"'] \\ E \arrow[r, "m"'] & B \end{tikzcd}$$

and by the dual of lemma 1.4.2 *(v)* it follows that $k \in \mathcal{E}$. Thus, all $\mathcal{E}$-factorizations of maps factor by an $\mathcal{E}$-map through the maximal factorization.

(iii) $\Rightarrow$ ***(v)*** We set $\mathcal{M}$ to be the maps which when $\mathcal{E}$-maximally factorized have their $\mathcal{E}$-factor an isomorphism.

[F.1] It suffices to prove that $\mathcal{M}$ contain all isomorphisms. Let $g : A \longrightarrow B$ be an isomorphism then g has a maximal factorization as $g = eg'$ but this means, as g is an $\mathcal{E}$-map, we have

$$\begin{tikzcd} A \arrow[rr, "g"] \arrow[d, equal] & & B \arrow[d, equal] \arrow[dl, dotted, "k"] \\ A \arrow[r, "e"'] & E \arrow[r, "m"'] & B \end{tikzcd}$$

a k such that $km = 1_B$ and $gk = e$. But this means that mk is the identity as it mediates between the maximal factorization and itself:

$$\begin{tikzcd} A \arrow[rr, "e"] \arrow[d, equal] & & E \arrow[d, "m"] \arrow[dl, dotted, "mk"'] \\ A \arrow[r, "e"'] & E \arrow[r, "m"'] & B \end{tikzcd}$$

This, in turn, makes m and k isomorphisms and thus $e = gm^{-1}$ is an isomorphism. Thus, **[F.1]** holds.

[F.2] Suppose now that m_1 and m_2 are $\mathcal{M}$-maps consider the maximal $\mathcal{E}$-factorization of m_1m_2, $m_1m_2 = em$, then we have unique k_2 and k_1 such that:

$$\begin{tikzcd} A \arrow[rr, "e"] \arrow[d, equal] & & E \arrow[d, "m"] \arrow[dl, dotted, "k_2"] \arrow[dll, dotted, "k_1"'] \\ A \arrow[r, "m_1"'] & B \arrow[r, "m_2"'] & C \end{tikzcd}$$

induced by the maximal factorizations of $m_2 = 1_M m_2$ and $m_1 1_A m_1$ respectively. However, it is easy to see that k_1 is not only a retract but also a section using the fact that $e \in \mathcal{E}$ and $m_1m_2 = em$ is a maximal factorization and

$$\begin{tikzcd} A. \arrow[rr, "e"] \arrow[d, equal] & & E \arrow[d, "m"] \arrow[dl, dotted, "k_1e"'] \\ A \arrow[r, "e"'] & E \arrow[r, "m"'] & C \end{tikzcd}$$

commutes it follows that $k_1e = 1_E$.

Thus, e is an isomorphism and $\mathcal{M}$ is closed to composition. This means **[F.2]** holds.

[F.3] The fact that maps can be factorized $\mathcal{E}$-maximally as $f = ef'$ is assumed: it suffices to show that f' is in $\mathcal{M}$: that is if f' is maximally factorized as $f' = e'f''$ then e' must be as isomorphism. To see this consider the two cross maps obtained from:

$$\begin{tikzcd} A \arrow[rr, "e"] \arrow[d, "e"'] & & E \arrow[d, "f'"] \arrow[dl, dotted, "v"'] \\ E \arrow[r, "e'"'] & E' \arrow[r, "f''"'] & B \end{tikzcd} \qquad \begin{tikzcd} A \arrow[r, "e"] \arrow[d, equal] & E \arrow[r, "e'"] & E' \arrow[d, "f''"] \arrow[dl, dotted, "w"] \\ A \arrow[r, "e"'] & E \arrow[r, "f'"'] & B \end{tikzcd}$$

where note that $ee' \in \mathcal{E}$ as it is closed to composition. It is clear that w is inverse to v and so they are isomorphisms. However, clearly also in the first diagram e' also serves as a cross map, whence $v = e'$ and e' is an isomorphism.

[F.4] The property of being a maximal $\mathcal{E}$-factorization for $\mathcal{M}$-maps amounts now to the requirement of orthogonality.

(v) $\Rightarrow$ ***(i)*** Immediate.

□

We shall say that a system of maps $\mathcal{M}$ is **left factor closed** if whenever $fg \in \mathcal{M}$ then $f \in \mathcal{M}$. Similarly a system of maps $\mathcal{E}$ is **right factor closed** if whenever fg in $\mathcal{E}$ then $g \in \mathcal{E}$. Thus, the monics and the sections are left factor closed while the epics and retractions are right factor closed. It is important to realize that the $\mathcal{M}$-maps of a factorization system need not necessarily be left-factor closed. Indeed this is a rather special condition upon which we shall improve when we have discussed limits. In the meantime we have the following rather technical observations for arbitrary factorization systems:

Lemma 1.4.6 *If $(\mathcal{E}, \mathcal{M})$ is a factorization system then*

(i) *If $fg \in \mathcal{M}$ and $g \in \mathcal{M}$ then $f \in \mathcal{M}$,*

(ii) *If $fg \in \mathcal{E}$ and $f \in \mathcal{E}$ then $g \in \mathcal{E}$,*

(iii) *$\mathcal{M}$ is left factor closed if and only if all $\mathcal{E}$-maps which are sections are isomorphisms (i.e. $\mathcal{M}$ contains all sections);*

(iv) *$\mathcal{E}$ is right factor closed if and only if all $\mathcal{M}$ maps which are retractions are isomorphisms (i.e. $\mathcal{E}$ contains all retractions).*

Proof: The first two parts are dual and follow immediately from lemma 1.4.2 *(v)*.

It suffices to prove one of the latter two statements as they are dual. Suppose therefore that $\mathcal{M}$ is left factor closed and $s \in \mathcal{E}$ has a right inverse then this right inverse makes s a left factor of an $\mathcal{M}$-map (the identity) and so s is in $\mathcal{M}$ as well. Thus, s is an isomorphism.

Conversely, if $fg \in \mathcal{M}$ and yet $f = em$ where $e \in \mathcal{E}$ then the square:

$$\begin{tikzcd} A \arrow[r, "e"] \arrow[d, equal] & B \arrow[d, "mg"] \arrow[dl, dotted, "k"] \\ A \arrow[r, "fg"'] & D \end{tikzcd}$$

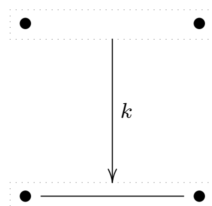
has a unique cross map making e a section and, thus, by assumption e is an isomorphism. Whence $em = f \in \mathcal{M}$. $\square$

Consider the following examples:

1. Set has a factorization system of maps into surjections followed by injections. This factorization clearly has the $\mathcal{M}$-maps left factor closed and the $\mathcal{E}$-maps right factor closed.

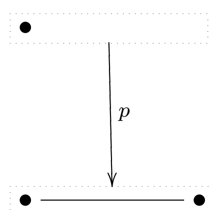
2. Consider the category of interference graphs, Intf. It is clear that maps between interference graphs can be factorized into epimorphisms generated by equivalence relations on the finite sets which are disjoint from the interference. Two equivalence classes interfere precisely when any two of their elements interfere. And monics, that is maps which are injective on the underlying sets.

   However, we are interested in some other more unusual factorization systems. Let $\mathbf{2}$ be the graph consisting of two isolated points and S_2 the graph of two points, x and y, which interfere $x \bowtie y$. There is an obvious map $k : \mathbf{2} \longrightarrow S_2$ which picks out the two points of S_2.

   The maps, $\mathcal{R}$, orthogonal to k are exactly the maps f which reflect the interference (that is such that $f(x) \bowtie f(y) \Rightarrow x \bowtie y$). Maps which are orthogonal to interference reflecting maps, S, are those which are isomorphism on the underlying sets but which increase the amount of interference or separation. It is not hard to see that $(\mathcal{S}, \mathcal{R})$ provides a factorization system on Intf.

3. Consider again the category of interference graphs, Intf. Let p be the map which picks out a point in S_2:

   Then the maps orthogonal to this are the maps which preserve the degree of each node. We shall call this class $\mathcal{C}$, the class of **cover** maps. The degree of a node in an interference graph is the number of other points which are connected to it.

   Can you characterize the maps which are left orthogonal to the cover maps? Do these maps and covers provide a factorization system on the category of infinite interference graphs?

1.4.3 Exercises

(1) Verify that the maps which are left orthogonal to the injections in Set are the surjections.

(2) Describe what a maximal orthogonal pair is in a poset.

(3) Let $\mathsf{Set}^{\mathbf{2}}$ be the category whose objects are maps of Set and whose maps are commutative squares. Thus, $(h_1, h_2) : f \longrightarrow g$ is a map if and only if

$$\begin{tikzcd} A \arrow[r, "f"] \arrow[d, "h_1"'] & B \arrow[d, "f"] \\ C \arrow[r, "g"'] & D \end{tikzcd}$$

Composition is by pasting of squares. Let $\mathbb{D}_0$ be the collection of maps which have h_1 an isomorphism then prove that $\mathbb{D}_0$ is a maximal right orthogonal class. Describe the left orthogonal class $(\mathbb{D}_0)_\perp$. Show that this gives a factorization system.

(4) A category is *completely orthogonal* in case every square has a unique cross-map. Show that groupies are always completely orthogonal. Show that the path category on a directed graph is completely orthogonal.

(5) (The Zappa-Szép product) Sometimes a category $\mathbb{X}$ has a *strict factorization system* this consist two classes of maps $\mathcal{E}$ and $\mathcal{M}$ which are closed to composition and are such that each map has a unique factorization as an $\mathcal{E}$-map followed by a $\mathcal{M}$-map. We may regard the two classes, $\mathcal{E}$ and $\mathcal{M}$, as two categories in their own right on the same set of objects. The composition in the whole category can then be viewed as producing a pair (f, g) with $f \in \mathcal{E}$ and $g \in calM$ so that $\partial_1(f) = \partial_0(g)$. The composition can then be broken down: $(f, g)(f', g') = (f(g \triangleleft f'), (g \triangleright f')g)$ where if $f : A \longrightarrow B$, $g : B \longrightarrow C$, $f' : C \longrightarrow D$, $g' : D \longrightarrow E$ then $g \triangleleft f' : B \longrightarrow g \bowtie f'$ $g \triangleright f' : g \bowtie f' \longrightarrow D$ so that we have:

$$\begin{tikzcd} & C \arrow[dr, "f'"] & \\ B \arrow[ur, "g"] \arrow[dr, "g \triangleleft f'"'] & & D \\ & g \bowtie f' \arrow[ur, "g \triangleright f'"'] & \end{tikzcd}$$

There are thus three induced functions starting at the arrow set $ME = \{(g, f) | g \in \mathcal{M}, f \in \mathcal{E}, \partial_1(g) = \partial_0 f\}$

$$\_ \triangleleft \_ : ME \longrightarrow \mathcal{E} \qquad \_ \triangleright \_ : ME \longrightarrow \mathcal{M} \qquad \_ \bowtie \_ : ME \longrightarrow \mathbb{X}_{\mathsf{obj}}$$

these must satisfy:

[ZS.1] $\partial_0(g \triangleleft f) = \partial_0(g)$, $\partial_1(g \triangleleft f) = g \bowtie f$, $\partial_0(g \triangleright f) = g \bowtie f$, and $\partial_1(g \triangleright f) = \partial_1(f)$;

[ZS.2] $(gh) \triangleleft f = g \triangleleft (h \triangleleft f)$ and $1 \triangleleft f = f$;

[ZS.3] $(gh) \triangleright f = (g \triangleright (h \triangleleft f))(h \triangleright f)$ and $1 \triangleright f = 1$;

[ZS.4] $g \triangleright (uv) = (g \triangleright u) \triangleright v$ and $g \triangleright 1 = g$

[ZS.5] $g \triangleleft (uv) = (g \triangleleft u)((g \triangleright u) \triangleleft v)$ and $g \triangleright 1 = 1$.

The data for a Zappa-Szép product of $\mathcal{E}$ and $\mathcal{M}$, two categories on the same objects, is given by three functions as above which satisfy the conditions [**ZS.1**]–[**ZS.5**].

(a) Show that a strict factorization can be equivalently given by having two orthogonal classes $\mathcal{E} \perp \mathcal{M}$, so that $\mathcal{E}$ is right factor closed ($f, fg \in \mathcal{E}$ implies $g \in \mathcal{E}$), $\mathcal{M}$ is left factor closed ($g, fg \in \mathcal{M}$ implies $f \in \mathcal{M}$) and every map can be uniquely factorized.

(b) Show that any category with a strict factorization produces the data for a Zappa-Szép product on the $\mathcal{E}$-maps and the $\mathcal{M}$-maps.

(c) Show that given the data for a Zappa-Szép product one can define a composition on pairs $(f,g)(f',g') := (f(g \triangleleft f'), (g \triangleright f')g)$ to produce a category, $\mathcal{E} \bowtie \mathcal{M}$, with a strict factorization with $\mathcal{E}$-maps being of the form $(g,1)$ and $\mathcal{M}$-maps being of the form $(1,f)$. This category is called the Zappa-Szép product of $\mathcal{E}$ and $\mathcal{M}$.

(d) Show the induced Zappa-Szép product of the two classes of a strict factorization system produces a category which is isomorphic (in a factorization preserving manner) to the original category. Conversely show that the induced Zappa-Szép data of the two classes of the strict factorization of a Zappa-Szép product is isomorphic to the original Zappa-Szép data.

1.5 Functors and natural transformations

It is natural to consider maps between categories, otherwise known as functors, and it is no surprise to discover that these organize themselves into a category Cat. What is a little more surprising is that there are also maps which can be defined between functors. This means that Cat is in fact a category which is enriched in categories.

1.5.1 Functors

A **functor** is a map of categories $F : \mathbb{C}_1 \longrightarrow \mathbb{C}_2$ which consists of a map F_0 of the objects and a map F_1 of the maps (we shall consistently drop these subscripts when the intended domain is clear) such that

- $D_0(F_1(f)) = F_0(D_0(f))$ and $D_1(F_1(f)) = F_0(D_1(f))$;
- $F_1(1_A) = 1_{F_0(A)}$;
- $F_1(fg) = F_1(f)F_1(g)$;

Clearly every category has an identity functor and the composition of functors is associative so that the following is immediate.

Lemma 1.5.1 *Categories and functors form a category* Cat*.*

If $F : \mathbb{C} \longrightarrow \mathbb{D}$ is a functor then $F^{\mathrm{op}} : \mathbb{C}^{\mathrm{op}} \longrightarrow \mathbb{D}^{\mathrm{op}}$ is a functor. A functor from the opposite of a category, $\mathbb{C}^{\mathrm{op}}$, is often called a **contravariant** functor from $\mathbb{C}$ as the maps get reversed (the *contra-* prefix). An ordinary functor (with domain $\mathbb{C}$) is sometimes called a **covariant** functor to emphasize that there is no twisting of maps involved.

A diagram in a category is a collection of arrows and objects satisfying certain composability relations. Suppose that P is a property involving a relationship between the maps and objects of a diagram. Examples of such properties, for the diagram consisting of a single map, that we have met so far are

- $P(f : A \longrightarrow B) = \exists g : B \longrightarrow A.fg = 1_A$, which says f is a section,
- $P(f : A \longrightarrow B) = \forall h, h' : X \longrightarrow A.hf = h'f \Rightarrow h = h'$ which says f is monic.

We shall say a functor F **preserves** a property P if whenever P holds of the arrows $f_1, .., f_n$ and objects $A_1, .., A_n$ then P holds of the arrows $F(f_1), .., F(f_n)$ and objects $F(A_1), .., F(A_n)$ in the codomain category.

Thus, a functor always preserves sections (and isomorphisms) but does not, in general, preserve monics (or epics). While functors may not preserve a property we are often interested in restriction attention to those functors which do preserve that property. Thus, while funcors do not in general preserve monics but we may well be interested in functors which do preserve monics.

We say that a functor **reflects** a property P of a diagram if whenever the property holds of the image under F of the diagram then the property must have held for the original arrows. (Notice that as functors do not in general even reflect composability, the requirement that the composability relations already hold in the domain is now important).

Thus, for example while functors always preserves isomorphisms they do not in general reflect isomorphisms.

The enriched view of a category gives another important view of a functor $F : \mathbb{C} \longrightarrow \mathbb{D}$ as being provided by a family of maps

$$F_{AB} : \mathbb{C}(A,B) \longrightarrow \mathbb{D}(F_0(A), F_0(B))$$

which must satisfy the following two diagrams:

$$\begin{tikzcd}
\mathbb{C}(A,B) \times \mathbb{C}(B,C) \arrow[r, "m_{ABC}"] \arrow[d, "F_{AB} \times F_{BC}"'] & \mathbb{C}(A,C) \arrow[d, "F_{AC}"] \\
\mathbb{D}(F_0(A), F0(B)) \times \mathbb{D}(F_0(B), F_0(C)) \arrow[r, "m_{F_0(A)F_0(B)F_0(C)}"'] & \mathbb{D}(F_0(A), f_0(C))
\end{tikzcd}$$

$$\begin{tikzcd}
1 \arrow[r, "1_A"] \arrow[dr, "1_{F_0(A)}"'] & \mathbb{C}(A,A) \arrow[d, "F_{AA}"] \\
& \mathbb{D}(F_0(A), F_0(A))
\end{tikzcd}$$

Using this point of view the idea of a **faithful** functor can be easily explained: it is a functor all of whose maps F_{AB} are injective. Similarly a **full** functor is one all of whose maps F_{AB} are surjective.

Notice that a category has a faithful functor to the final category (one object one map) if and only if it is a preorder! As we shall see in a moment, every (small) category has a full functor to a preorder.

Congruences

Given a category $\mathbb{C}$ a **congruence** is a given by an equivalence relation on each hom-set satisfying $f \sim g$ and then $hfk \sim hgk$. Given a congruence we may form a new category with the same objects $\mathbb{C}/\sim$ whose hom-sets are the $\sim$–equivalence classes: $\mathbb{C}(A,B)/\sim_{AB}$. This is called the **quotient category** for the congruence $\sim$:

Objects: Those of $\mathbb{C}$;

Maps: Equivalence classes of maps under $\sim$ we may write these as $[f] : A \longrightarrow B$ where $f : A \longrightarrow B$ is a representative member of the equivalence class.

Identities: $[1_A] : A \longrightarrow A$;

Composition: $[f][g] = [fg]$.

Now it is not immediate that this is indeed a category. It is clear that the identities have the correct properties but what is not clear is that composition so defined is even a function. Specifically it is not clear that if we have $[f_1] = [f_2]$ and $[g_1] = [g_2]$ (in other words $f_1 \sim f_2$ and $g_1 \sim g_2$) that necessarily $f_1 g_1$ and $f_2 g_2$ are even related! Of course this is exactly where we must use the special property of a congruence, here is the argument:

$$\dfrac{\dfrac{f_1 \sim f_2}{f_1 g_1 \sim f_2 g_1}\ \text{Compose } 1\_g_1 \qquad \dfrac{g_1 \sim g_2}{f_2 g_1 \sim f_2 g_2}\ \text{Compose } f_2\_1}{f_1 g_1 \sim f_2 g_2}\ \text{Transitive}$$

There is an obvious functor $Q_\sim : \mathbb{C} \longrightarrow \mathbb{C}/\sim$ which is the identity on objects and carries a map f to its $\sim$–equivalence class $[f]$. This is clearly always a full functor.

There is an important way in which conguences arise:

Lemma 1.5.2 *If $F : \mathbb{X} \longrightarrow \mathbb{Y}$ is a functor then the relation on parallel arrows $f \sim_F g \Leftrightarrow F(f) = F(g)$ is a congruence. Furthermore F can be factorized as $Q_{\sim_F} F'$ where $Q_{\sim_F}$ full and the identity on objects and F' is faithful.*

PROOF: We must show that $hfk \sim_F hgk$ but $F(hfk) = F(h)f(f)F(k) = F(h)F(g)F(k) = F(hgk)$.

We may define a functor $F' : \mathbb{X}/\sim_F \longrightarrow \mathbb{Y}$ by $F'([f]) = F(f)$. This is clearly well-defined and, furthermore, a faithful functor as if $F'([f]) = F'([g])$ then $F(f) = F(g)$. $\square$

It is reasonable to wonder whether this is a factorization system on functors: and, indeed, this is the case. The two classes of functors are

- $\mathcal{Q}$ the quotient functors, these are full functors which are an isomorphisms on the objects (that is F such that $F_0 : \mathbb{X}_0 \longrightarrow \mathbb{Y}_0$ is an isomorphism and each $F_1^{A,B} : \mathbb{X}(A,B) \longrightarrow \mathbb{Y}(F(A),F(B))$ is a surjection). Note that all quotient functors are, as functors, epic as both their object part and their map parts are surjections.

- $\mathcal{F}$ the faithful functors (that is F such that each $F_1^{A,B} : \mathbb{X}(A,B) \longrightarrow \mathbb{Y}(F(A),F(B))$ is an injection).

To show this is a factorization system first note that we have already obtained the factorization of a functors in lemma 1.5.2: it remains to check the other conditions of part *(i)* of proposition 1.4.5. Note that composing either class with functors which are isomorphisms keeps one in the class so the only problem is to show that quotient functors are orthogonal to faithful functors. For this consider:

$$\begin{tikzcd}
\mathbb{X} \arrow[r, "Q"] \arrow[d, "G"'] & \mathbb{Y} \arrow[d, "H"] \arrow[dl, dotted, "V"'] \\
\mathbb{X}' \arrow[r, "F"'] & \mathbb{Y}'
\end{tikzcd}$$

where Q is a quotent functor, F is faithful, and we need to define V. On objects we are forced to define V as $V(A) = G(Q^{-1}(A))$ as Q is an isomorphism on objects. Consider, therefore, $V(f)$ for $f : A \longrightarrow B$: as Q is full $f = Q(x)$ for some $x : Q^{-1}(A) \longrightarrow Q^{-1}(B)$: we then define $V(f) = G(x)$. This is well-defined as if $Q(x) = Q(y)$ for some parallel arrow y then $G(y) = G(x)$ as $F(G(x)) = F(G(y))$ and F is faithful. It is then easy to check that V so defined is a functor: it preserves identities as $V(1_A) = G(1_{Q^{-1}(A)})$, it preserves composition as, suppose $f = Q(x)$ and $g = Q(y)$ then $V(fg) = G(xy) = G(x)G(y) = V(f)V(g)$. This immediately makes the triangle commute $Q; V = G$ commute by the way we defined V. The other triangle also must commute as clearly $Q; H = G; F = Q; V; F$ and as Q is epic as a functor it follows $H = V; F$.

This gives:

Proposition 1.5.3 *For the category of categories $(\mathcal{Q}, \mathcal{F})$, the quotiont functors and the faithful functors, give a factorization system.*

Note that faithful functors need not, as functors, be monic by any means.

Functors of many variables

Often functors will have their domain a product of two or more categories. This means the functor will have more than one argument. Consider a functor $F : \mathbb{A} \times \mathbb{B} \longrightarrow \mathbb{C}$ then, clearly, fixing either argument of F at an object of $\mathbb{A}$ or $\mathbb{B}$ induces functors:

$$F_A = F(A, \_) : \mathbb{B} \longrightarrow \mathbb{C} \quad \text{and} \quad F_B = F(\_, B) : \mathbb{A} \longrightarrow \mathbb{C}$$

conversely given a family of such functors F_A and F_B we may reconstruct F:

Proposition 1.5.4 *To give a functor $F : \mathbb{A} \times \mathbb{B} \longrightarrow \mathbb{C}$ is to have a family of functors $F_A : \mathbb{B} \longrightarrow \mathbb{C}$ for each $A \in \mathbb{A}$ and $F_B : \mathbb{A} \longrightarrow C$ for each $B \in \mathbb{B}$ such that $F_A(B) = F_B(A)$ and for any $g : A \longrightarrow A' \in \mathbb{A}$ and $f : B \longrightarrow B' \in \mathbb{B}$ the following equality $F_A(f)F_{B'}(g) = F_B(g)F_{A'}(f)$.*

PROOF: Clearly if F is a functor then we have F_A and F_B with these properties the content of the result is in the reverse direction. Define $F(f, g) = F_A(f)F_{B'}(g) = F_B(g)F_{A'}(f)$: we must show that this is a functor.

First notice that this definition ensures that the identity maps are preserved. For composition we have

$$\begin{aligned} F(f,g)F(f',g') &= F_{A_1}(g)F_{B_2}(f)F_{A_2}(g')F_{B_3}(f') \\ &= F_{A_1}(g)F_{A_1}(g')F_{B_3}(f)F_{B_3}(f') \\ &= F_{A_1}(gg')F_{B_3}(ff') \\ &= F(ff', gg'). \end{aligned}$$

□

1.5.2 Natural transformations

Given two functors $F, G : \mathbb{C} \longrightarrow \mathbb{D}$ a **natural transformation** (or just a transformation) $\alpha : F \Rightarrow G$ is a family of maps $\alpha_C : F(C) \longrightarrow G(C)$, indexed by the objects of $\mathbb{C}$, in $\mathbb{D}$ such that for every map $f : C_1 \longrightarrow C_2$ in $\mathbb{D}$ the following diagram commutes:

$$\begin{tikzcd} F(C_1) \arrow[r, "F(f)"] \arrow[d, "\alpha_{C_1}"'] & F(C_2) \arrow[d, "\alpha_{C_2}"] \\ G(C_1) \arrow[r, "G(f)"'] & G(C_2) \end{tikzcd}$$

Our first observation is that this means that $\mathsf{Cat}(\mathbb{C}, \mathbb{D})$, which often written $\mathbb{D}^{\mathbb{C}}$, can be given the structure of a category. Ultimately this means that Cat is a Cat-enriched category – these are also known as **2-categories**.

Proposition 1.5.5 $\mathsf{Cat}(\mathbb{C},\mathbb{D})$ *is a category with objects functors and maps natural transformations.*

PROOF: First notice that every functor has an identity transformation given by $1_{F(A)} : F(A) \to F(A)$. To compose natural transformations we simply define $(\alpha\beta)_A = \alpha_A\beta_A$: if this composition works then it is associative. It remains to check that the above requirement on the composite transformation holds: this can be seen by pasting the transformation squares together.

$$\begin{tikzcd} F(C_1) \arrow[r, "F(f)"] \arrow[d, "\alpha_{C_1}"'] & F(C_2) \arrow[d, "\alpha_{C_2}"] \\ G(C_1) \arrow[r, "G(f)"'] \arrow[d, "\beta_{C_1}"'] & G(C_2) \arrow[d, "\beta_{C_1}"] \\ H(C_1) \arrow[r, "H(f)"'] & H(C_2) \end{tikzcd}$$

□

In order to conveniently manipulate functors and natural transformations it is useful to develop the 2-categorical view of them. As there are many levels of activity in a general 2-category it is useful to introduce a special notation: the objects of a 2-category (in Cat these are categories) are often called **0-cells** while the maps (or functors) are called **1-cells** and the transformations between the 1-cells are called **2-cells**.

In this enriched view of Cat there are two sorts of composition: the composition of natural transformations, $\alpha\beta$, written by juxtaposition and the composition of functors written which we shall write with a semicolon, $F;G$.

Proposition 1.5.6 Cat *is a* Cat*–enriched category.*

PROOF: The main difficulty is to prove that functor composition

$$\_;\_ : \mathsf{Cat}(\mathbb{A},\mathbb{B}) \times \mathsf{Cat}(\mathbb{B},\mathbb{C}) \longrightarrow \mathsf{Cat}(\mathbb{A},\mathbb{C})$$

is a functor of two arguments. To show this we describe the functors $F;\_$ and $\_;G$ and then argue, using proposition 1.5.4, that from these we can reconstruct a functor in two arguments.

If $\alpha : F \Rightarrow G$ then define $(H;\alpha)_A = \alpha_{H(A)}$: this is a natural transformation as α is. Furthermore, it is clear that this is a functorial assignment as it clearly preserves composition.

Define $(\alpha;K)_A = K(\alpha_A)$: this is a natural transformation also as α is and K is a functor. This also is clearly a functorial assignment.

In establishing that this is a bifunctor it remains to check that $(\alpha;H)(G;\beta) = (F;\beta)(\alpha;K)$ this is provided by the following commutative diagram which relies on the naturality of β:

$$\begin{tikzcd} H(F(A)) \arrow[r, "H(\alpha_A)"] \arrow[d, "\beta_{F(A)}"'] & H(G(A)) \arrow[d, "\beta_{G(A)}"] \\ K(F(A)) \arrow[r, "K(\alpha_A)"'] & K(F(A)) \end{tikzcd}$$

It remains to check that the associative and identity laws hold for the 1-cell composition $\_;\_$. Associativity of these functors applied to natural transformations is given by the following commutative diagram where each face commutes by naturality (we already know that functor composition is associative):

$$\begin{tikzcd}
J(H(F(A))) \arrow[rrr, "J(H(\alpha_A))"] \arrow[ddd, "J(\beta_{F(A)})"'] \arrow[dr, "\gamma_{H(F(A))}"] & & & J(H(G(A))) \arrow[ddd, "J(\beta_{G(A)})"] \arrow[dl, "\gamma_{H(G(A))}"'] \\
& L(H(F(A))) \arrow[r, "L(H(\alpha_A))"'] \arrow[d, "L(\beta_{F(A)})"'] & L(H(G(A))) \arrow[d, "L(\beta_{G(A)})"] & \\
& L(K(F(A))) \arrow[r, "L(K(\alpha_A))"] & L(K(G(A))) & \\
J(K(F(A))) \arrow[rrr, "J(K(\alpha_A))"'] \arrow[ur, "\gamma_{K(F(A))}"'] & & & J(K(G(A))) \arrow[ul, "\gamma_{K(G(A))}"]
\end{tikzcd}$$

It is straightforward to check the identity laws: we take the product with the final category and the functor which picks out the identity functor and identity transformation. We must check that this composition has no effect which is immediate. □

As the double semicolon is an enriched composition we may apply it to the natural transformations as well as to the functors so that if $\alpha : F \longrightarrow G$ and $\alpha' : I \longrightarrow J$ then

$$\alpha; \alpha' : F; I \longrightarrow G; J$$

As a consequence of the enrichment one will then have the equation, which is called the **interchange law** which is a direct consequence of the fact that the composition is a functor of two arguments.

$$(\alpha; \beta)(\alpha'; \beta') = (\alpha\alpha'); (\beta\beta')$$

In order to represent these compositions we shall use pasting diagrams: these diagrams are very useful as interchange and associativity laws become graphical equalities. This means that one can immediately "see" that two composites are equal with a minimal amount of manipulation. Below is the representation of the interchange law:

$$\begin{tikzcd}
\mathbb{A} \arrow[rr, bend left=60, "F"] \arrow[rr, "G" description] \arrow[rr, bend right=60, "H"'] \arrow[r, phantom, shift left=3ex, "\Downarrow\alpha"] \arrow[r, phantom, shift right=3ex, "\Downarrow\beta"] & & \mathbb{B} \arrow[rr, bend left=60, "I"] \arrow[rr, "J" description] \arrow[rr, bend right=60, "K"'] \arrow[r, phantom, shift left=3ex, "\Downarrow\alpha'"] \arrow[r, phantom, shift right=3ex, "\Downarrow\beta'"] & & \mathbb{C}
\end{tikzcd}$$

Pasting diagrams may be translated back into 2-categorical notation but not in an unambiguous way as there are many ways of representing a given pasting diagram as in the 2-categorical "combinator" notation we have introduced. For example the above interchange can also be written:

$$
\begin{aligned}
&(\alpha; J)(\beta; \alpha')(H; \beta') \\
&(\alpha; (\alpha'\beta'))(\beta; K) \\
&(F; \alpha')(\alpha; \beta')(\beta; K) \\
&....
\end{aligned}
$$

It is important to realize that not every pasting diagram is legal. Here are a description of the conditions to ensure that it is legal:

(a) The diagram must be planar with nodes labeled by objects (0-cells) arrows by functors (1-cells) and regions by (2-cells);

(b) There must be a start node (object or 0-cell) and an end node (object or 0-cell);

(c) The maps (functors or 1-cell) must form an acyclic directed graph;

(d) Every map (functor or 1-cell) must be on a path from the start node to the end node, which we will call a trip.

(e) For the transformations (2-cells) there must be a starting trip (path from start node to end node) and an ending trip.

(f) Every arrow must be either on the end trip or in the domain of exactly one transformation (but not both);

(g) Every arrow must be either on the start trip or in the codomain of exactly one transformation (but not both).

Notice that the conditions (b), (c), and (d) must already be true of commuting diagrams.

These systems are actually familiar in computer science: they occur as in formal languages. The rules of a context free language may be seen as 2-cells. The system above allows us a notation for describing the derivations in the language.

1.5.3 The Yoneda lemma

Many categories we shall consider have a natural enrichment in Set and for such categories there are some rather natural functors, the so-called hom-functors. The Yoneda lemma characterizes the natural transformations between hom-functors completely and has the consequence that any category $\mathbb{C}$ embeds fully and faithfully in $\mathsf{Set}^{\mathbb{C}^{\mathrm{op}}}$.

Give a Set-enriched category $\mathbb{C}$ we may consider for each object $A \in \mathbb{C}$ the functor:

$$\mathbb{C}(\_, A) : \mathbb{C}^{\mathrm{op}} \longrightarrow \mathsf{Set}; X \mapsto \mathbb{C}(X, A)$$

and if $f : X \longrightarrow Y$ then $\mathbb{C}(f, A) : \mathbb{C}(Y, A) \longrightarrow \mathbb{C}(X, A); g \mapsto fg$. That this is a functor is not hard to check. Yoneda made the following crucial observation:

Lemma 1.5.7 *The natural transformations* $\alpha : \mathbb{C}(,A) \Rightarrow F : \mathbb{C}^{\mathrm{op}} \longrightarrow \mathsf{Set}$ *are in bijective correspondence to the elements of* $F(A)$.

PROOF: The trick to to set up a correspondence. If α is such a natural transformation define $\widehat{\alpha} = \alpha_A(1_A) \in F(A)$. Conversely, if $x \in F(A)$ define $\mathcal{Y}(x)_X(g) = F(g)(x)$. Now it is immediate that $\widehat{(\mathcal{Y}(x))} = \mathcal{Y}(x)_A(1_A) = F(1_A)(x) = x$. For the reverse we have

$$\begin{aligned}
\mathcal{Y}(\widehat{\alpha})_X(g) &= F(g)(\widehat{\alpha}) \\
&= F(g)(\alpha_A(1_A)) \\
&= \alpha_X(\mathbb{C}(g,A)(1_A)) \\
&= \alpha_X(g)
\end{aligned}$$

So provided $\mathcal{Y}(x)$ *is* a natural transformation the proof will be complete! To show this we must verify that

$$\begin{tikzcd}
\mathbb{C}(Y,A) \arrow[r, "\mathbb{C}(f,A)"] \arrow[d, "\mathcal{Y}(x)_Y"'] & \mathbb{C}(X,A) \arrow[d, "\mathcal{Y}(x)_X"] \\
F(Y) \arrow[r, "F(f)"'] & F(X)
\end{tikzcd}$$

To show this is the case we take an element $g \in \mathbb{C}(Y,A)$ and follow it round the diagram each way:

$$\begin{aligned}
\mathcal{Y}(x)_X(\mathbb{C}(f,A)(g)) &= \mathcal{Y}(x)_X(fg) \\
&= F(fg)(x) \\
&= F(f)(F(g)(x)) \\
&= F(f)(\mathcal{Y}(x)_Y(g)).
\end{aligned}$$

□

The Yoneda lemma has the following important consequence:

Corollary 1.5.8 *There is a full and faithful embedding* $\mathcal{Y} : \mathbb{C} \longrightarrow \mathsf{Set}^{\mathbb{C}^{\mathrm{op}}}$ *which has* $\mathcal{Y}(A) = \mathbb{C}(\_,A)$ *and* $\mathcal{Y}(f) = \mathbb{C}(f,A)$ *as defined above.*

PROOF: A natural transformation $\alpha : \mathbb{C}(\_,A) \Rightarrow \mathbb{C}(\_,B)$ is $\mathcal{Y}(\alpha_A(1_A))$ where $\alpha_A(1_A)$ is some map $f : A \longrightarrow B$. So that this gives a bijection between $\mathbb{C}(A,B)$ and $\mathsf{Set}^{\mathbb{C}^{\mathrm{op}}}(\mathbb{C}(\_,A), \mathbb{C}(\_,B))$.

It remains to check that this gives a functor: in particular we must show $\mathcal{Y}(1_A) = 1_{\mathbb{C}(,A)}$ and $\mathcal{Y}(f)\mathcal{Y}(g) = \mathcal{Y}(fg)$. For the former we have:

$$\mathcal{Y}(1_A)_X(g) = \mathbb{C}(g,A)(1_A) = g.$$

While for the latter we have:

$$\mathcal{Y}(fg)_X(k) = \mathbb{C}(k,C)(fg) = kfg = \mathcal{Y}(g)_X(kf) = \mathcal{Y}(g)_X(\mathcal{Y}(f)_X(k)) = (\mathcal{Y}(f)\mathcal{Y}(g))_X(k).$$

□

1.5.4 Exercises

(1) Prove that functors do not in general preserve monics or epics but that they do preserve sections and retractions.

(2) Prove that $\_ \times A : \mathsf{Set} \longrightarrow Sets; X \mapsto X \times A$ is a functor and the family of projections $\pi_1 : X \times A \longrightarrow X$ give a natural transformation.

(3) Prove that list : $\mathsf{Set} \longrightarrow \mathsf{Set}$, the process of forming lists, is functorial. Prove that the process of reversing a list is an endo-natural transformation on the list functor.

(4) Show that the only functor of a category $\mathbb{C}$ into the final category $\mathbf{1}$ is faithful if and only if $\mathbb{C}$ is a preorder.

(5) Show that directed graphs and directed graph homomorphisms are exactly the functor category $\mathsf{Set}^{\mathbf{2}}$ where $\mathbf{2}$ is the category:

$$\begin{tikzcd} A \arrow[loop left, "1_A"] \arrow[r, shift left, "a_0"] \arrow[r, shift right, "a_1"'] & B \arrow[loop right, "1_B"] \end{tikzcd}$$

in which the composition structure is forced.

(6) Construct a finite category whose functors into sets are "reflective symmetric graphs." That is graphs with an specified loop ι_a at each node a and for each arrow $a \xrightarrow{f} b$ a specified arrow $b \xrightarrow{\widehat{f}} a$ such that $\widehat{\widehat{f}} = f$ and $\widehat{\iota_a} = \iota_a$.

(7) A forest (F, β, α) is a graded set $\beta : F \longrightarrow \mathbb{N}$ together with an action

$$\alpha : \{(n, x) \in \mathbb{N} \times F | n \leq \beta(x)\} \longrightarrow F$$

such that $\beta(\alpha(n, x)) = n$, $\alpha(n, \alpha(m, x)) = \alpha(n, x)$, and $\alpha(\beta(x), x) = x$. A morphism of forests $f : (F, \beta, \alpha) \longrightarrow (F', \beta', \alpha')$ is a set maps $f : F \longrightarrow F'$ such that $\beta'(f(x)) = \beta(x)$ and $\alpha'(n, f(x)) = f(\alpha(n, x))$.

Prove that forests form a category Forest.

Provide a functor $H : \mathsf{Forest} \longrightarrow \mathsf{Set}^{\mathbb{N}^{\mathrm{op}}}$ where $\mathbb{N}$ is regarded as a poset with respect to the usual ordering.

(8) Write down a 2-categorical combinator form for the following 2-cell

$$\begin{tikzcd} & B \arrow[dr, "G"] \arrow[rrrd, bend left, "L"] & & \arrow[dl, phantom, "\Swarrow\beta"] & \\ A \arrow[ur, "F"] \arrow[dr, "H"'] & \arrow[d, phantom, "\Downarrow\alpha"] & D \arrow[r, "N"] \arrow[d, phantom, "\Downarrow\gamma"] & E \arrow[r, "O"'] & F \\ & C \arrow[ur, "K"'] \arrow[urr, bend right, "M"'] & & & \end{tikzcd}$$

(9) If $F : \mathbb{C} \longrightarrow \mathbb{D}$ show that the relation on the hom-sets induced by $f \sim g$ if and only if $F(f) = F(g)$ is a congruence.

(10) Prove carefully that $\mathbb{C}/\sim$ is a category whenever $\sim$ is a congruence and that $Q_\sim$ is a functor as advertised.

(11) Show that every functor may be factorized into a full (identity on objects) functor followed by faithful functor. (Hint: you can use the notion of a congruence.)

(12) Given any category $\mathbb{C}$ prove that if $\alpha : F \Rightarrow G : \mathbb{C} \longrightarrow \mathbb{D}$ is a natural transformation then there is a natural transformation

$$\mathsf{Split}(\alpha) : \mathsf{Split}(F) \Rightarrow \mathsf{Split}(G) : \mathsf{Split}(\mathbb{C}) \longrightarrow \mathsf{Split}(\mathbb{D}).$$

(13) If M is a monoid an M-set is given by a set S together with an action $\alpha : M \times S \longrightarrow S$ such that $\alpha(1, s) = s$ and $\alpha(m_1m_2, s) = \alpha(m_1, \alpha(m_2, s))$. A homomorphism $f : (S, \alpha) \longrightarrow (S'.\alpha')$ is a map $f : S \longrightarrow S'$ between the sets such that $f(\alpha(s, m)) = \alpha'(f(s'), m)$.

Show that M-sets form a category $M - \mathsf{Set}$. Prove that there is an isomorphism of categories $V : M - \mathsf{Set} \longrightarrow \mathsf{Set}^{M^{\mathrm{op}}}$.

An M-set (S, α) is freely generate by an element $s \in S$ in case given a choice of element $y \in S'$ for any other M-set (S', α') there is a unique homomorpism of M-sets $y^* : (S, \alpha) \longrightarrow (S', \alpha')$ such that $y^*(x) = x$. Prove that such an M-set exists. (Hint: use the Yoneda lemma!)

Chapter 2

Adjoints and Monad

The purpose of this chapter is to introduce one of the most important concepts of category theory: an adjunction. We have coupled this with the introduction of limits and colimits because these, of course, provide immediate examples of adjunctions and, through the adjoint functor theorems, are linked to limits and colimits.

However, we start with a brief discussion of some very basic constructions of categories. We have already met one very basic construction: the construction of the functor category $\mathbb{D}^{\mathbb{C}}$. This, however, is just one of many useful constructions.

2.1 Basic constructions on categories

Perhaps *the* most basic construction on categories (and many other structures) is the formation of the product of two categories. Recall that the product is formed by taking the cartesian product of both the objects and the maps and defining the composition coordinate-wise.

Slightly less usual is the the "coproduct" of two categories. Rather like directed graphs one can simply disjointly union two categories together. Thus, as in graph theory, we can talk about connected components of a category. For example, every finite category can always be decomposed uniquely into a sum of connected components.

Most categories we shall consider will be connected and, in fact, most of the constructions we will consider preserve connectedness. Notice that, in particular, the product of two connected categories is always guaranteed to be connected.

2.1.1 Slice categories

If $\mathbb{C}$ is any category and $X \in \mathbb{C}$ we may form the **slice category** $\mathbb{C}/X$. This has the following structure:

Objects: Maps of $\mathbb{C}$ to X, $f : C \longrightarrow X$;

Maps: Triples $(f_1, g, f_2) : f_1 \longrightarrow f_2$ which are commutative triangles:

$$\begin{tikzcd} C_1 \arrow[rr, "f"] \arrow[dr, "f_1"'] & & C_2 \arrow[dl, "f_2"] \\ & X & \end{tikzcd}$$

Identities: $(f, 1_C, f) : f \longrightarrow f$.

Composition: $(f_1, g, f_2)(f_2, h, f_3) = (f_1, gh, f_3)$ which is well-defined as $ghf_3 = gf_2 = f_1$.

Now one must, in fact, check that this is a category which amounts to checking that we have a composition which satisfies the required axioms. However, this structure is directly inherited from $\mathbb{C}$ so the proof is straightforward.

Slice categories will become important when we discuss type theories and fibrations. They are important to us immediately as they are a warm up for a more general construction. Before considering this more general construction it is, however, worth noting some peculiar things about the slice categories over Set:

$$\begin{array}{rcl} \mathsf{Set}/0 & \cong & \mathbf{1} \\ \mathsf{Set}/1 & \cong & \mathsf{Set} \\ \mathsf{Set}/2 & \cong & \mathsf{Set} \times \mathsf{Set} \\ \mathsf{Set}/3 & \cong & \mathsf{Set} \times \mathsf{Set} \times \mathsf{Set} \\ & \ldots & \end{array}$$

In fact, we may regard Set/I as the I-indexed product of the category of sets with itself. This is because a map $C \longrightarrow I$ in Set is "the same thing" as the I-indexed collection of sets $(C_i = \{c \in C | f(c) = i\})_{i \in I}$. A map in the slice category $k : f \longrightarrow f'$ can then be viewed as an indexed collection of maps $(k_i : C_i \longrightarrow C'_i)$.

This view of a slice category is rather special to sets. However, the properties of these slice categories of Set have also been the inspiration for trying to capture abstractly the special properties of Set.

2.1.2 Comma categories

Let $F : \mathbb{A} \longrightarrow \mathbb{C}$ and $G : \mathbb{B} \longrightarrow \mathbb{C}$ be functors then we may form the **comma category**[1] of F over G which is denoted F/G as follows:

Objects: Triples $(A, F(A) \xrightarrow{f} G(B), B)$;

Maps: Quadruples $((A, f, B), a : A \longrightarrow A', b : B \longrightarrow B', (A', f', B'))$ where (A, f, B) and (A', f', B') are objects as above and a and b are maps which render

$$\begin{tikzcd} F(A) \arrow[r, "f"] \arrow[d, "F(a)"'] & G(B) \arrow[d, "G(b)"] \\ F(A') \arrow[r, "f'"'] & G(B') \end{tikzcd}$$

commutative.

Identities: $((A, f, B), 1_A, 1_B, (A, f, B))$;

[1]This is a bit of a misnomer considering the notation chosen here! The original notation for this construction involved a comma.

Composition: With $((A_1, f_1, B_1), a, b, (A_2, f_2, B_2))((A_2, f_2, B_2), a', b', (A_3, f_3, B_3))$ defined to be $((A_1, f_1, B_1), aa', bb', (A_3, f_3, B_3))$ where the required commutativity is provided by:

$$\begin{tikzcd} F(A_1) \arrow[r, "f_1"] \arrow[d, "F(a)"'] & G(B_1) \arrow[d, "G(b)"] \\ F(A_2) \arrow[r, "f_2"'] \arrow[d, "F(a')"'] & G(B_2) \arrow[d, "G(b')"] \\ F(A_3) \arrow[r, "f_3"'] & G(B_3) \end{tikzcd}$$

Again the fact that this is a category must be checked although it follows easily from the fact that F and G are functors and that $\mathbb{A}$ and $\mathbb{B}$ are categories.

Let us first observe that, indeed, the slice category construction is a special case of this construction. First notice that an object in a category corresponds precisely to a functor from the final category $\mathbf{1}$. Thus $\mathbb{C}/A$ can be read as the comma category using the identity functor on $\mathbb{C}$, $\mathbb{C} = 1_{\mathbb{C}} : \mathbb{C} \longrightarrow \mathbb{C}$ and the functor from the final category $A : \mathbf{1} \longrightarrow \mathbb{C}$. It is easy now to see that this comma category is just the slice category.

The comma category F/G has some obvious associated functors

$$\Pi_0 : F/G \longrightarrow \mathbb{A}; ((A, f, B), a, b, (A', f', B')) \mapsto a$$

$$\Pi_1 : F/G \longrightarrow \mathbb{B}; ((A, f, B), a, b, (A', f', B')) \mapsto b$$

There is also a canonical natural transformation

$$\alpha : \Pi_0; F \longrightarrow \Pi_1; G$$

where $\alpha_{(A,f,B)} = f : F(A) \longrightarrow G(B)$. Thus, the comma category gives a pasting square:

$$\begin{tikzcd} F/G \arrow[r, "\Pi_0"] \arrow[d, "\Pi_1"'] & \mathbb{A} \arrow[d, "F"] \arrow[dl, Rightarrow, "\alpha"] \\ \mathbb{B} \arrow[r, "G"'] & \mathbb{C} \end{tikzcd}$$

This construction is actually a "weighted limit"(as are inserters) and has some rather special properties which are described in the exercises.

2.1.3 Inserters

If $F, G : \mathbb{X} \longrightarrow \mathbb{Y}$ are functors then we may form the **inserter category** of F over G which is denoted $F//G$ as the category:

Objects: Pairs $(A, F(A) \xrightarrow{f} G(A))$;

Maps: Triples $((A, f), a : A \longrightarrow A', (A', f'))$ where (A, f) and (A', f') are objects as above and a is a maps which render

$$\begin{tikzcd} F(A) \arrow[r, "f"] \arrow[d, "F(a)"'] & G(A) \arrow[d, "G(a)"] \\ F(A') \arrow[r, "f'"'] & G(A') \end{tikzcd}$$

commutative.

Identities: $((A, f), 1_A, (A, f))$;

Composition: $((A_1, f_1), a, (A_2, f_2))((A_2, f_2), a', (A_3, f_3)) = ((A_1, f_1), aa', (A_3, f_3))$ where the required commutativity is provided by:

$$\begin{tikzcd} F(A_1) \arrow[r, "f_1"] \arrow[d, "F(a)"'] & G(A_1) \arrow[d, "G(a)"] \\ F(A_2) \arrow[r, "f_2"'] \arrow[d, "F(a')"'] & G(A_2) \arrow[d, "G(a')"] \\ F(A_3) \arrow[r, "f_3"'] & G(A_3) \end{tikzcd}$$

Clearly there is a functor $U : F//G \longrightarrow \mathbb{X}$ and it is easy to see that there is a natural transformation $\alpha : U; F \Rightarrow U; G$ whose value at (A, f) is $f : F(A) \longrightarrow G(A)$.

Inserters are useful in the description of inductive and coinductive datatypes. Examples of inductive datatypes include the natural numbers, lists, binary trees, rose trees, etc. Coinductive datatypes include streams, conumbers, possibly infinite lists, and simple objects (in the object oriented sense). The datatype for an endo functor F can be defined by adjunction from an inserters of the form $F//\mathbb{X}$ (or $\mathbb{X}//F$) which may also be viewed as the category of algebras (respectively coalgebras) for that functor. This topic is discussed in the next chapter.

2.1.4 Exercises

(1) The signature of an algebra is a specification, for each operation symbol, of the number of arguments it has. The category of signatures of algebras is $Sets/\mathbb{N}$. Justify this claim and explain the morphisms of signatures.

(2) A comulti-graph is part of the structure of a context free grammar. The category of comultigraphs (in Set) is $\mathsf{Set}/\langle\mathsf{Set}, \mathrm{list}\rangle$. Explain this claim and explain the morphisms of these multigraphs.

(3) The category of "magma" – sets with one binary operation is called a magma (a term due Bourbaki)– is $\langle\mathsf{Set}, \mathsf{Set}\rangle//\mathsf{Set}$. Justify this claim and describe the morphisms of magma.

(4) There is an obvious functor $U : \mathbb{C}/X \longrightarrow \mathbb{C}; (x, f, x') \mapsto f$ prove that this functor

   (a) Reflect isomorphisms

(b) Preserve monics

(c) Reflects monics

(Harder!) Does the functor reflect and preserve epics: provide a counter-example.

(5) If $f : X \longrightarrow Y$ prove that there is a functor

$$f_* : \mathbb{C}/X \longrightarrow \mathbb{C}/Y; (x, g, x') \mapsto (xf, g, x'f).$$

Does this functor preserve and reflect monics and isomorphisms?

(6) The (1-dimensional) special property of the comma category: suppose $\beta : H; F \Rightarrow J; G : \mathbb{X} \longrightarrow \mathbb{Z}$ then prove there is a unique functor $K : \mathbb{X} \longrightarrow F/G$ such that $K; \Pi_0 = H$ and $K; \Pi_1 = J$ and that $\beta = K; \alpha$ where α is the canonical 2-cell $\alpha : \Pi_0; F \Rightarrow \Pi_1; G$.

(7) The (1-dimensional) special property of the inserter: Suppose $\beta : H; F \Rightarrow H; G : \mathbb{X} \longrightarrow \mathbb{Z}$ then show that there is a unique functor $K : \mathbb{X} \longrightarrow F//G$ such that $K; U = H$ and $\beta = K; \alpha$ where $\alpha : U; F \Rightarrow U;: G$ is the canonical 2-cell.

(8) Suppose $f : P_1 \longrightarrow P$ and $g : P_2 \longrightarrow P$ are order preserving maps of posets describe the poset f/g. Now if $P_1 = P_2$ describe $f//g$.

2.2 Adjoints

In this section we introduce one of the most important concepts of basic category theory: the notion of an adjunction. This notion occurs all over mathematics and plays an absolutely fundamental role in understanding how abstract settings, such as a programming languages, are constructed.

2.2.1 The universal property

Let $G : \mathbb{Y} \longrightarrow \mathbb{X}$ be a functor and X an object of $\mathbb{X}$, then an object $U \in \mathbb{Y}$ together with a map $\eta_X : X \longrightarrow G(U)$ is a **universal pair** for the functor G at the object X if for any $f : X \longrightarrow G(Y)$ there is a unique $f^\sharp : U \longrightarrow Y$ such that

$$\begin{tikzcd} X \arrow[r, "\eta_X"] \arrow[dr, "f"'] & G(U) \arrow[d, dotted, "G(f^\sharp)"] \\ & G(Y) \end{tikzcd}$$

commutes.

It is useful to have in mind a particular instance of this universal property. A nice example is as follows: let $\mathbb{X}$ be the category of directed graphs and $\mathbb{Y}$ the category of categories, let the functor G be the "underlying functor" which forgets the compositional structure of a category, that is regards a category as no more than the "underlying" directed graph. The map which takes a directed graph and embeds it into the graph underlying the path category as the singleton paths (paths of length one) has the universal property for this "underlying" functor.

Consider a map of directed graphs into the graph underlying a category, $h : G \longrightarrow U(\mathbb{C})$, we can extend it uniquely to a functor from the path category to the category as follows. Let $h^\sharp : \mathsf{Path}(G) \longrightarrow \mathbb{C}$ be defined on arrows by

$$h^\sharp(A, [a_1, .., a_n], B) = h(a_1)..h(a_n) : h(A) \longrightarrow h(B)$$

then it is easy to check that this is a functor and that is uniquely determined by h. Before proceeding, it is well worth doing the exercise to check that this does really work.

In the next section on limits and colimits we shall be discussing another important situation in which these universal properties hold. However, we shall continue here to develop the theory of these universal properties first. The reader who needs more concrete motivation may like to start reading the section on limits and colimits in parallel.

We first make the simple but crucial observation:

Lemma 2.2.1 *If (η_X, U) and (η'_X, U') are universal at X for $G : \mathbb{Y} \longrightarrow \mathbb{X}$ then there is a unique isomorphism $\alpha : U \longrightarrow U'$ such that $\eta_X G(\alpha) = \eta'_X$.*

PROOF: We may define α as the unique map $(\eta'_X)^\sharp$:

$$\begin{tikzcd} X \arrow[r, "\eta_X"] \arrow[dr, "\eta'_X"'] & G(U) \arrow[d, dotted, "G(\alpha)"] \\ & G(U') \end{tikzcd}$$

and by swapping the role of U and U' we obtain a $\beta : U' \longrightarrow U$ using the universal property of (U', η'_X). However, now

$$\eta_X G(\alpha\beta) = \eta_X G(\alpha)G(\beta) = \eta'_X G(\beta) = \eta_X$$

so $\alpha\beta = \eta_X^\sharp$ and $\eta_X^\sharp = 1_U$ so $\alpha\beta = 1_U$ and by a similar argument $\beta\alpha = 1_{U'}$. □

Thus a universal pair (U, η) for a functor G at an object X is determined upto a unique isomorphism.

We may also use the Yoneda lemma to re-express this property: if $F : \mathbb{Y} \longrightarrow \mathsf{Set}$ and there is a natural isomorphism $\alpha : \mathbb{Y}(U, \_) \rightarrow F$ then F is said to be **representable** with **universal element** $\alpha_U(1_U) \in F(U)$. In this case we have:

$$\alpha : \mathbb{Y}(U, \_) \Rightarrow \mathbb{X}(X, G(\_)) : \mathbb{Y} \longrightarrow \mathsf{Set}$$

where $\alpha_Y(h) = \eta_X G(h)$ and the inverse maps sends f to $f^\sharp$. Thus the universal property above can be re-expressed by saying that $\mathbb{X}(X, G(\_))$ is a representable functor with universal element $\eta_X \in \mathbb{X}(X, G(U))$. This reminds us that specifying the map η_X actually suffices to determine the situation.

Now it is certainly not always the case that, for a functor G there will be universal pairs $(F(X), \eta_X)$ at each object X, however, if this is the case we have:

Proposition 2.2.2 *Let $G : \mathbb{Y} \longrightarrow \mathbb{X}$ be a functor such that for each $X \in \mathbb{X}$ there is a universal pair $(F(X), \eta_X)$ then:*

- *F is a functor with $F(g) = (g\eta)^\sharp$;*
- *$\eta_X : X \longrightarrow G(F(X))$ is a natural transformation;*
- *$\epsilon_Y = 1^\sharp_{G(Y)} : F(G(Y)) \longrightarrow Y$ is a natural transformation;*
- *The* **triangle equalities** *$\eta_{G(Y)}G(\epsilon_Y) = 1_{G(Y)}$ and $F(\eta_X)\epsilon_{F(X)} = 1_{F(X)}$ hold.*

Conversely given functors F and G with transformations η and ϵ, as above, which satisfy the triangle identities, then each $(F(X), \eta_X)$ is universal for G at X.

PROOF: We start by verifying that F so defined is a functor it will then be immediate that η is a natural transformation. We must verify that F preserves identities, which is the observation that $(\eta_X)^\sharp = 1$, and that $F(f)F(g) = F(fg)$. For the latter we have that, as

$$\eta_X G((f\eta_Y)^\sharp(g\eta_Z)^\sharp) = \eta_X G((f\eta_Y)^\sharp)G((g\eta_Z)^\sharp) = f\eta_Y G((g\eta_Z)^\sharp) = fg\eta_Z$$

that $(f\eta_Y)^\sharp(g\eta_Z)^\sharp = (fg\eta_Z)^\sharp$.

It remains to prove that $\epsilon_Y = 1^\sharp_{G(Y)}$ is a natural transformation and that the second triangle equality holds – the first is immediate from the definition of ϵ. For the naturality of ϵ we have for a map $f : Y \longrightarrow Y'$ that

$$G(f) = \eta_{G(Y)}G((1_{G(Y)})^\sharp)G(f) = \eta_{G(Y)}G(\epsilon_Y)G(f)$$

so that $G(f)^\sharp = \epsilon_Y f$. Similarly,

$$G(f) = G(f)\eta_{G(Y')}G((1_{G(Y')})^\sharp) = \eta_{G(Y)}G(F(G(f))G((1_{G(Y')})^\sharp) = \eta_{G(Y)}G(F(G(f))G(\epsilon_{Y'}))$$

so that $G(f)^\sharp = G(F(G(f)))\epsilon_{Y'}$. Thus, ϵ is natural.

Finally, for the second triangle equality we have:

$$\eta_X G(F(\eta_X)\epsilon_{F(X)}) = \eta_X \eta_{F(X)}\epsilon_{F(X)}) = \eta_X$$

so that $F(\eta_X)\epsilon_{F(X)} = 1_{F(X)} = \eta^\sharp_X$.

For the converse, suppose we have (F, G, η, ϵ) and the triangle equalities then, given $f : X \longrightarrow G(Y)$ we can set $f^\sharp = F(f)\epsilon_Y$ then, using the naturality of η and the first triangle equality we have:

$$\eta_X G(f^\sharp) = \eta_X G(F(f)\epsilon_Y) = f\eta_{G(X)}\epsilon_Y = f.$$

To show uniqueness we suppose $\eta_X G(h) = f$ then

$$h = F(\eta_X)\epsilon_{F(X)}h = F(\eta_X)F(G(h))\epsilon_Y = F(\eta_X G(h))\epsilon_Y = F(f)\epsilon_Y = f^\sharp.$$

□

We say that F is **left adjoint** to G (equally G is **right adjoint** to F: notice that this is also the dual statement) in case, as in the proposition there are natural transformations η (called the **unit** of the adjunction) and ϵ (called the **counit** of the adjunction) satisfying the triangle equalities. We write this situation as:

$$(\eta, \epsilon) : F \vdash G : \mathbb{X} \longrightarrow \mathbb{Y}.$$

Recall there is a dual concept: let $F : \mathbb{X} \longrightarrow \mathbb{Y}$ be a functor and Y an object of $\mathbb{Y}$, then an object $V \in \mathbb{X}$ together with a map $\epsilon_Y : F(V) \longrightarrow Y$ is a **couniversal pair** for the functor F at the object Y if for any $g : F(X) \longrightarrow Y$ there is a unique $g^\flat : X \longrightarrow V$ such that

$$\begin{tikzcd} F(X) \arrow[r, "g"] \arrow[d, dotted, "F(g^\flat)"'] & Y \\ F(V) \arrow[ur, "\epsilon_Y"'] & \end{tikzcd}$$

commutes.

When F has for each $Y \in \mathbb{Y}$ a couniversal pair $(G(Y), \epsilon_Y)$ then this also gives rise to an adjunction but this time the constructed functor G is right adjoint to F. In terms of universal elements we now have a natural isomorphism:

$$\alpha : \mathbb{X}(\_, V) \Rightarrow \mathbb{Y}(F(\_), Y) : \mathbb{X}^{\mathrm{op}} \longrightarrow \mathsf{Set}.$$

which is determined by the **couniversal element** ϵ_Y. The fact, that this is a dual concept is apparent in this formulation as we have replaced $\mathbb{X}$ by $\mathbb{X}^{\mathrm{op}}$.

2.2.2 Basic properties of adjoints

There is the following characterization of an adjoints which here we state for Set–enriched categories. The result is true more generally for categories enriched elsewhere (e.g. Cat-enriched categories for example).

Theorem 2.2.3 *The following are equivalent for* Set*–enriched categories and functors* $F : \mathbb{X} \longrightarrow \mathbb{Y}$ *and* $G : \mathbb{Y} \longrightarrow \mathbb{X}$*:*

(i) *An adjoint* $(\eta, \epsilon) : F \dashv G : \mathbb{X} \longrightarrow \mathbb{Y}$*;*

(ii) *Two combinators* $(\_)^\flat$ *and* $(\_)^\sharp$ *where*

$$(g : F(A) \longrightarrow B)^\flat : A \longrightarrow G(B) \quad \text{and} \quad (f : A \longrightarrow G(B))^\sharp : F(A) \longrightarrow B$$

such that

- $(f^\sharp)^\flat = f$ *and* $(g^\flat)^\sharp = g$,
- $(F(h) f k)^\flat = h f^\flat G(k)$ *and* $(k' g G(h'))^\sharp = F(k') g^\sharp h'$*;*

(iii) *A natural isomorphism* $(\_)^\flat : \mathbb{Y}(F(\_), \_) \Rightarrow \mathbb{X}(\_, G(\_)) : \mathbb{X}^{\mathrm{op}} \times \mathbb{Y} \longrightarrow \mathsf{Set}$*;*

(iv) *An isomorphism of categories* $(\_)^\sharp : \mathbb{X}/G \longrightarrow \mathbb{F}/\mathbb{Y}$ *such that* $(\_)^\sharp; \Pi_0 = \Pi_0$ *and* $(\_)^\sharp; \Pi_1 = \Pi_1$*;*

(v) *For each* $X \in \mathbb{X}$ *there is a pair* $(F(X), \eta_X)$ *which is universal for* G *at* X *such that* $F(h) = (h\eta)^\sharp$*;*

(vi) *For each* $Y \in \mathbb{Y}$ *there is a pair* $(G(Y), \epsilon_Y)$ *which is couniversal for* F *at* Y *such that* $G(k) = (\epsilon k)^\flat$*.*

PROOF: We already know that the universal properties (and couniversal properties for that matter) are equivalent to giving adjoints and it is clear that *(iv)* and *(v)* are dual. Thus it remains to prove the following:

(i) $\Rightarrow$ ***(ii)*** : Given the above adjunction we may define $h^\flat = \eta G(h)$ and $g^\sharp = F(g)\epsilon$. Now these are inverse as

$$(h^\flat)^\sharp = (\eta G(h))^\sharp = F(\eta G(h))\epsilon = F(\eta)\epsilon h = h$$

and the argument for $(f^\sharp)^\flat = f$ is dual. Now notice that

$$(F(h)fk)^\flat = \eta G(F(h)fk) = h\eta G(f)G(k) = hf^\flat G(k)$$

and the other identity follows by the dual argument.

(ii) $\Leftrightarrow$ ***(iii)*** : The "sharp" and "flat" now are going to be viewed as natural isomorphisms; the announced properties of sharp and flat are exactly saying that regarded as transformations they are natural!

(ii) $\Leftrightarrow$ ***(iii)*** : The sharp and flat are now going to be regarded as functors. The commuting square:

$$\begin{tikzcd} X \arrow[r, "h"] \arrow[d, "f"'] & G(Y) \arrow[d, "G(g)"] \\ X' \arrow[r, "h'"'] & G(Y') \end{tikzcd}$$

is a map $(f, g) : h \longrightarrow h'$ in $\mathbb{X}/G$ under the "sharp" combinator it is carried to the square:

$$\begin{tikzcd} F(X) \arrow[r, "h^\sharp"] \arrow[d, "F(f)"'] & Y \arrow[d, "g"] \\ F(X') \arrow[r, "(h')^\sharp"'] & Y' \end{tikzcd}$$

which commutes as $F(f)(h')^\sharp = (fh')^\sharp = (hG(g))^\sharp = h^\sharp g$. Thus, "sharp" does indeed give a functor which clearly has an inverse and, furthermore, this functor commutes with the projections from the two slice categories as required.

Conversely if we are given such an isomorphism of categories then by "sharping" the following two diagrams:

$$\begin{tikzcd} X \arrow[r, "fh"] \arrow[d, "f"'] & G(Y) \arrow[d, equal] \\ X' \arrow[r, "h"'] & G(Y) \end{tikzcd} \qquad \begin{tikzcd} X' \arrow[r, "h"] \arrow[d, equal] & G(Y) \arrow[d, "G(k)"] \\ X' \arrow[r, "hG(k)"'] & G(Y') \end{tikzcd}$$

we obtain:

$$\begin{tikzcd} F(X) \arrow[r, "(fh)^\sharp"] \arrow[d, "F(f)"'] & Y \arrow[d, equal] \\ F(X') \arrow[r, "h^\sharp"'] & Y \end{tikzcd} \qquad \begin{tikzcd} F(X') \arrow[r, "h^\sharp"] \arrow[d, equal] & Y \arrow[d, "k"] \\ F(X') \arrow[r, "(hG(k))^\sharp"'] & Y' \end{tikzcd}$$

Which gives the combinator identity

$$(k'gG(h'))^\sharp = F(k)(gG(h))^\sharp = F(k')g^\sharp h'$$

in two steps. The other required identity for the "flat" functor follows dually.

(ii) $\Rightarrow$ ***(iv)*** : We must now recover the universal property from the "sharp" and "flat" combinators: set $\eta_X = (1_{F(X)})^\flat$ then given $f : X \longrightarrow G(Y)$ we have show

$$(1_{F(X)})^\flat G(f^\sharp) = (1_{F(X)} f^\sharp)^\flat = (f^\sharp)^\flat = f$$

and it remains to show that $f^\sharp$ is unique. So suppose $f = (1_{F(X)})^\flat G(g)$ then

$$f^\sharp = ((1_{F(X)})^\flat G(g))^\sharp = ((1_{F(X)} g)^\flat)^\sharp = (g^\flat)^\sharp = g.$$

Finally we note

$$(f\eta)^\sharp = (f(1_{F(X')})^\flat)^\sharp = F(f)((1_{F(X')})^\flat)^\sharp = F(f)1_{F(X')} = F(f)$$

So that the definition of F agrees with that given by the universal property.

$\square$

We shall often write the two-way transformation as a two-way logical inference:

$$\frac{X \xrightarrow{f = g^\flat} G(Y)}{F(X) \xrightarrow[g = f^\sharp]{} Y} \; F \dashv G$$

This allows us to move from maps $h : F(X) \longrightarrow Y$ to $h^\flat : X \longrightarrow G(Y)$ and back $f : X \longrightarrow G(Y)$ to $f^\sharp : F(X) \longrightarrow Y$ where these moves are mutually inverse and natural. This style of handling adjoints as we shall see shortly allows a convenient way of reasoning about the effect of adjoints.

Left adjoints preserve universal pairs and dually, although we do not formally state it, right adjoints preserve couniversal pairs:

Proposition 2.2.4 *If $G : \mathbb{Y} \longrightarrow \mathbb{X}$ and $(\eta', \epsilon') : H \dashv K : \mathbb{Y} \longrightarrow \mathbb{Z}$ and (U, η') is a universal pair for then $(H(U), \eta_U \eta')$ is a universal pair for $K; G$.*

PROOF: Suppose $f : X \longrightarrow G(K(Z))$ then we obtain a unique $f^{\sharp'} : U \longrightarrow K(Z)$ now we may use the universal property of the adjunction to give a unique $(f^{\sharp'})^\sharp : H(U) \longrightarrow Z$ such that $\eta' G(f^{\sharp'})^\sharp) = f$.

$$\frac{\dfrac{X \xrightarrow{f} G(K(Z))}{U \xrightarrow{f^{\sharp'}} K(Z)} \; \text{Universal}}{H(U) \xrightarrow[(f^{\sharp'})^\sharp]{} Z} \; H \dashv K$$

□

This, in particular, means that if G has a left adjoint then $K;G$ will have a left adjoint whose unit is $\eta_X \eta'_{F(X)}$. This gives:

Corollary 2.2.5 *If F and G are left (right) adjoints then $F;G$ is a left (right) adjoint.*

PROOF: While the above remarks establish that this observation is correct we may also see this very directly by using the inferences:

$$\cfrac{\cfrac{X \xrightarrow{f} G(K(Z))}{F(X) \xrightarrow{f^{\sharp'}} K(Z)}\; F \dashv G}{H(F(X)) \xrightarrow[(f^{\sharp'})^{\sharp}]{} Z}\; H \dashv K$$

□

2.2.3 Reflections, coreflections and equivalences

As a preliminary for our next topic we observe:

Proposition 2.2.6 *If $(\eta, \epsilon) : F \dashv G : \mathbb{X} \longrightarrow \mathbb{Y}$ is an adjunction then*

(i) *Each η_X is monic if and only if F is faithful;*

(ii) *Each ϵ_Y is epic if and only if G is faithful;*

(iii) *Each η_X is a retraction if and only if F is full;*

(iv) *Each ϵ_Y is a section if and only if G is full;*

(v) *Each η_X is an isomorphism if and only if F is full and faithful;*

(vi) *Each ϵ_Y is an isomorphism if and only if G is full and faithful.*

PROOF:

(i) If $F(f) = F(g)$ where $f, g : X \longrightarrow X'$ then $f\eta = \eta G(F(f)) = \eta G(F(g)) = g\eta$ and so, as η is monic $f = g$. This shows F is faithful.

Conversely if F is faithful then and $f\eta = g\eta$ then $F(f) = (f\eta)^{\sharp} = (g\eta)^{\sharp} = F(g)$.

(ii) This is dual to *(i)*.

(iii) This is dual to *(iv)*.

(iv) Suppose that ν_Y is right inverse to ϵ_Y (note that ν is *not* assumed to be natural) then given $f : G(Y) \longrightarrow G(Y')$ we obtain $\nu_Y F(f)\epsilon_{Y'} : Y \longrightarrow Y'$ and observe:

$$G(\nu_Y F(f)\epsilon_{Y'}) = (\epsilon_Y \nu_Y F(f)\epsilon_{Y'})^\flat = (F(f)(1_{G(Y')})^\sharp)^\flat = f((1_{G(Y')})^\sharp)^\flat = f$$

showing that G is full.

Conversely suppose G is full then there is a $k : X \longrightarrow F(G(X))$ such that $\eta_{G(X)} = G(k) : G(X) \longrightarrow G(F(G(X)))$ but then $\epsilon k = F(G(k))\epsilon = F(\eta_{G(X)}\epsilon = 1_{F(G(X))}$ and so ϵ is a section.

(v) Combine *(i)* and *(iii)*.

(vi) Combine *(ii)* and *(iv)* or notice this is dual to *(v)*.

□

We shall say that a full and faithful functor $I : \mathbb{X}' \longrightarrow \mathbb{X}$ is a **reflection** in case I is a right adjoint. By the above proposition this means that if

$$(\eta, \epsilon) : F \dashv I : \mathbb{X} \longrightarrow \mathbb{X}'$$

is the adjunction then each ϵ_X is an isomorphism.

Dually we shall say that the full and faithful functor $I : \mathbb{X}' \longrightarrow \mathbb{X}$ is a **coreflection** in case I is a left adjoint. By the above proposition this means that if

$$(\eta, \epsilon) : I \dashv G : \mathbb{X}' \longrightarrow \mathbb{X}$$

is the adjunction then each η'_X is an isomorphism.

A full and faithful functor F is said to be an **equivalence** of categories in case it has an (either left or right) adjoint which is also full and faithful. This means that both the unit and counit of the adjunction are isomorphisms. Here is a slightly different and more useful formulation of an equivalence:

Proposition 2.2.7 *The following are equivalent:*

(i) $(\eta, \epsilon) : F \dashv G : \mathbb{X} \longrightarrow \mathbb{Y}$ *an equivalence of categories;*

(ii) *A full and faithful functor* $F : \mathbb{X} \longrightarrow \mathbb{Y}$ *and for each* $Y \in \mathbb{Y}$ *an* $X \in \mathbb{X}$ *and an isomorphism* $\alpha_Y : Y \longrightarrow F(X)$.

PROOF: Given the adjunction we may choose $X = G(Y)$ and $\alpha = \epsilon^{(-1)}$. The content is in the reverse direction. For the converse we check the couniversal property:

$$\begin{tikzcd} F(X) \arrow[d, dotted, "\alpha h"'] & Y \arrow[l, "\alpha^{(-1)}"'] \arrow[dl, "h"] \\ F(Z) & \end{tikzcd}$$

as F is full and faithful there is a unique $k : X \longrightarrow A$ such that $F(k) = \alpha h$. □

Now a left adjoint does not necessarily reflect isomorphisms but it is important to note that a reflection does (and dually for a coreflection) which fact is immediate as reflections are full and faithful.

Suppose now $F : \mathbb{X} \longrightarrow \mathbb{Y}$ is a reflection then consider the full subcategory $\mathbb{Y}'$ of $\mathbb{Y}$ determined by those objects which are isomorphic to some $F(X)$. A subcategory which contains all isomorphic "copies" of objects is often called a **wide** (or **replete**) subcategory. Thus, $\mathbb{Y}'$ is the wide subcategory generated by the image of F. Now clearly F restricts to $F' : \mathbb{X} \longrightarrow \mathbb{Y}'$ and by construction of $\mathbb{Y}'$ the functor F' is full and faithful and for every $Y \in \mathbb{Y}'$ there is an $F(X)$ which is isomorphic to it (so modulo the ability to choose the isomorphism) F is an equivalence of categories. However, now by composing adjoints we immediately have that $\mathbb{Y}' \subseteq \mathbb{Y}$ is a reflective subcategory.

2.2.4 Adjoints in 2-categories

There is another almost painless way of understanding the abstract properties of adjoints which involves looking at them in a Cat–enriched category setting, that is in a 2-category.

An adjunction in a 2-category is given by (F, G, η, ϵ) where $F : \mathbb{X} \longrightarrow \mathbb{Y}$ and $G : \mathbb{Y} \longrightarrow \mathbb{X}$ are 1-cells such that $\eta : 1_{\mathbb{X}} \Rightarrow F; G$ and $\epsilon : G; F \longrightarrow 1_{\mathbb{Y}}$ satisfying the triangle equalities $(G; \eta)(\epsilon; G) = 1_G$ and $(F; \epsilon)(\eta; F) = 1_F$. These identities can be drawn as pasting diagrams:

$$
\begin{tikzcd}
\mathbb{X} \arrow[rr, equal] \arrow[dr, "F"'] & \arrow[d, phantom, "\Downarrow\eta"] & \mathbb{X} \arrow[dr, "F"] \arrow[d, phantom, "\Downarrow\epsilon"] & \\
& \mathbb{Y} \arrow[ur, "G"] \arrow[rr, equal] & \, & \mathbb{Y}
\end{tikzcd}
\quad = \quad
\begin{tikzcd}
\mathbb{X} \arrow[dr, "F"] & \\
& \mathbb{Y}
\end{tikzcd}
$$

$$
\begin{tikzcd}
& \mathbb{X} \arrow[rr, equal] \arrow[dr, "F"] \arrow[d, phantom, "\Downarrow\epsilon"] & \arrow[d, phantom, "\Downarrow\eta"] & \mathbb{X} \\
\mathbb{Y} \arrow[ur, "G"] \arrow[rr, equal] & \, & \mathbb{Y} \arrow[ur, "G"'] &
\end{tikzcd}
\quad = \quad
\begin{tikzcd}
& \mathbb{Y} \\
\mathbb{Y} \arrow[ur, "G"'] &
\end{tikzcd}
$$

Using these diagrams one can often surprisingly easily establish the basic properties of adjunctions. For example suppose we wish to prove that adjunctions compose then we simply need to note that we can paste the adjunction diagrams together. Thus, suppose $(\eta, \epsilon) : F \dashv G : \mathbb{X} \longrightarrow \mathbb{Y}$ and $(\eta', \epsilon') : F' \dashv G' : \mathbb{Y} \longrightarrow \mathbb{Z}$ then we have:

$$
\begin{tikzcd}
\mathbb{X} \arrow[rrrr, equal] \arrow[dr, "F"'] & & \arrow[d, phantom, "\Downarrow\eta"] & & \mathbb{X} \arrow[dr, "F"] \arrow[d, phantom, "\Downarrow\epsilon"] & & \\
& \mathbb{Y} \arrow[rr, equal] \arrow[dr, "F'"'] & \arrow[d, phantom, "\Downarrow\eta'"] & \mathbb{Y} \arrow[ur, "G"] \arrow[rr, equal] & \arrow[d, phantom, "\Downarrow\epsilon'"] & \mathbb{Y} \arrow[dr, "F'"] & \\
& & \mathbb{Z}' \arrow[ur, "G"'] \arrow[rrrr, equal] & & \, & & \mathbb{Y}
\end{tikzcd}
\quad = \quad
\begin{tikzcd}
\mathbb{X} \arrow[dr, "F"'] & & \\
& \mathbb{Y} \arrow[dr, "F'"'] & \\
& & \mathbb{Z}
\end{tikzcd}
$$

which together with the dual diagram shows how to compose adjoints.

We shall now use these techniques to prove:

Proposition 2.2.8 *In any 2-category if $F \dashv G$ and $F \dashv G'$ if and only if G is isomorphic to G'.*

Proof: Supposing G and G' are right adjoints of F we must construct 2-cells $G \longrightarrow G'$ and $G' \longrightarrow G$. These are

$$
\begin{tikzcd}
 & \mathbb{X} \arrow[rr, equal] \arrow[dr, "F"] \arrow[d, phantom, "\Downarrow\epsilon"] & & \mathbb{X} \\
\mathbb{Y} \arrow[ur, "G"] \arrow[rr, equal] & \; & \mathbb{Y} \arrow[ur, "G'"'] \arrow[u, phantom, "\Downarrow\eta'"] &
\end{tikzcd}
\qquad
\begin{tikzcd}
 & \mathbb{X} \arrow[rr, equal] \arrow[dr, "F"] \arrow[d, phantom, "\Downarrow\epsilon'"] & & \mathbb{X} \\
\mathbb{Y} \arrow[ur, "G'"] \arrow[rr, equal] & \; & \mathbb{Y} \arrow[ur, "G"'] \arrow[u, phantom, "\Downarrow\eta"] &
\end{tikzcd}
$$

Pasting these together easily shows that they are isomorphisms. The converse we leave as an exercise. □

2.2.5 Exercises

(1) Show in detail that the underlying functor $U : \mathsf{Cat} \longrightarrow \mathsf{Graph}$ has a left adjoint given by the path category construction.

(2) Let $\mathbb{R}$ be the real numbers viewed as a category by using the usual ordering; the integers may also be regarded as a category, $\mathbb{Z}$, by using the usual ordering and the usual inclusion of the integers into the reals is a functor. Prove that this functor has a left and right adjoint.

(3) Let P be a poset. A **closure operator** on P is and order preserving mapping $J : P \longrightarrow P$ such that $x \leq J(x)$ and $J(J(x)) = J(x)$. Show that closure operators correspond precisely to coreflective subcategories of P.

(4) A **Galois connection** is a contravariant adjunction between posets. That is $f \dashv g : P_1^{\mathrm{op}} \longrightarrow P_2$.

   (a) Show that the relationship between classes of maps given by right orthogonal and left orthogonal form a Galois connection.

   (b) Show that $g(f(g(x))) = g(x)$ and $f(g(f(y))) = f(y)$ for any Galois connection and that the full subposets determined by $\{x | x = g(f(x))\} \subseteq P_1$ and $\{y | y = f(g(y))\} \subseteq P_2$ are (contra-)isomorphic.

(5) (Folklore) Call an adjunction $(\eta, \epsilon) : F \dashv G : \mathbb{X} \longrightarrow \mathbb{Y}$ a Galois adjunction if all of $\eta; F$, $F; \epsilon$, $G; \eta$, and $\epsilon; G$ are isomorphisms.

   (a) Prove that the assumption that any one of the above is an isomorphism forces all to be isomorphisms,

   (b) Show that such an adjunction (between ordinary categories) can be factorized in to the composite of reflection followed by a coreflection.

   This is the general form for the previous question on Galois connections.

(6) We may view Rel as a 2-category – more precisely as a poset enriched category – where the 2-structure is given by inclusions between relations. Prove that a relation is a left-adjoint if and only if it is a function. What is the right adjoint?

(7) Given an endo-relation R on a set X we may define the following operations on the powers set $\mathcal{P}(X)$:

$$
\begin{array}{rcl}
\Diamond\!\cdot\, \alpha & = & \{x | \exists y \in \alpha \cdot x \sim_R y\} \\
\cdot\!\Diamond\, \alpha & = & \{x | \exists y \in \alpha \cdot y \sim_R x\} \\
\Box\!\cdot\, \alpha & = & \{x | \forall y \in \alpha \cdot x \sim_R y\} \\
\cdot\!\Box \alpha & = & \{x | \forall y \in \alpha \cdot y \sim_R x\}
\end{array}
$$

Which operations (if any) are adjoints?

(8) Let $f : X \longrightarrow Y$ be any map of sets then prove that this induces a chain of adjoints $\exists_f \dashv f^* \dashv \forall_f : \mathcal{P}(X) \longrightarrow \mathcal{P}(Y)$ where

- $\mathcal{P}(X)$ and $\mathcal{P}(Y)$ are the powersets (set of all subsets) of X and Y respectively,
- $\exists_f : \mathcal{P}(X) \longrightarrow \mathcal{P}(Y); S \mapsto \{y \in Y | \exists x \in X \cdot f(x) = y \wedge x \in S\}$,
- $f^* : \mathcal{P}(Y) \longrightarrow \mathcal{P}(X); T \mapsto \{x \in X | f(x) \in T\}$,
- $\forall_f : \mathcal{P}(X) \longrightarrow \mathcal{P}(Y); S \mapsto \{y \in Y | \forall x \in X \cdot f(x) = y \Rightarrow x \in S\}$.

(9) Provide and alternative proof using the universal property that if $F \dashv G : \mathbb{X} \longrightarrow \mathbb{Y}$ and $F \dashv G'$ then G is naturally isomorphic to G'.

(10) Prove that the underlying functor $U : A \times \_//\mathsf{Set} \longrightarrow \mathsf{Set}$ has a left adjoint (hint: think lists).

(11) The category of posets has an obvious inclusion into the category of preorders. Prove that this is a reflection (hint: how do you turn a preorder into an order?).

(12) Show that in any 2-category given 2-cell isomorphisms $\alpha : 1_{\mathbb{X}} \longrightarrow F; G$ and $\beta : 1_{\mathbb{Y}} \longrightarrow G; F$ there is an adjoint equivalence between F and G.

2.3 Monads and comonads

An adjunction $(\eta, \epsilon) : F \vdash G : \mathbb{X} \longrightarrow \mathbb{Y}$ generates an endofunctor both on the category $\mathbb{X}$, $F; G : \mathbb{X} \longrightarrow \mathbb{X}$, and on the category $\mathbb{Y}$, $G; F : \mathbb{Y} \longrightarrow \mathbb{Y}$, together with certain natural transformations. The section studies these induced structures which are called, respectively, monads and comonads[2] and describes two important constructions associated with them: the Eilenberg-Moore construction and the Kleisli construction.

2.3.1 The definition

A **monad** $\mathbb{T} = (T, \eta, \mu)$ on a category $\mathbb{X}$ consists of an endofunctor $T : \mathbb{X} \longrightarrow \mathbb{X}$ and two natural transformations, the **unit** $\eta_X : X \longrightarrow T(X)$ and the **multiplication** $\mu : T(T(X)) \longrightarrow T(X)$ these

[2] As with many important concepts these have actually enjoyed a variety of names: they are still often called "triples". However, the term *monad* (popularized by MacLane) seems to have endured: this name suggests a relationship to monoids – and indeed there is!

must satisfy the following equations:

$$\begin{tikzcd} T(X) \arrow[r, "\eta_{T(X)}"] \arrow[dr, equal] & T(T(X)) \arrow[d, "\mu"] & T(X) \arrow[l, "T(\eta_X)"'] \arrow[dl, equal] \\ & T(X) & \end{tikzcd}$$

$$\begin{tikzcd} T(T(T(X))) \arrow[r, "\mu"] \arrow[d, "T(\mu)"'] & T(T(A)) \arrow[d, "\mu"] \\ T(T(X)) \arrow[r, "\mu"'] & T(X) \end{tikzcd}$$

A monad is, in fact, a monoid in the category of endofunctors where the "tensor product" is given by functor composition. The first two equations are therefore referred to as the unit equations while the last is called the associativity equation.

The dual concept is a **comonad** $\mathcal{S} = (S, \epsilon, \delta)$ where $\epsilon_X : S(X) \longrightarrow X$ and $\delta_X : S(X) \longrightarrow S(S(X))$ such that the dual of the above equations hold:

$$\begin{tikzcd} & S(X) \arrow[dl, equal] \arrow[d, "\delta_X"] \arrow[dr, equal] & \\ S(X) & S(S(X)) \arrow[l, "\epsilon_{S(X)}"] \arrow[r, "S(\epsilon_X)"'] & S(X) \end{tikzcd}$$

$$\begin{tikzcd} S(X) \arrow[r, "\delta_X"] \arrow[d, "\delta_X"] & S(S(X)) \arrow[d, "S(\delta_X)"] \\ S(S(X)) \arrow[r, "\delta_{S(X)}"'] & S(S(S(X))) \end{tikzcd}$$

A comonad is a comonoid in the category of endofunctors.

Our first observation is:

Proposition 2.3.1 *If $(\eta, \epsilon) : F \dashv G : \mathbb{X} \longrightarrow \mathbb{Y}$ is an adjoint then $\mathbb{T} = (F; G, \eta, F; \epsilon; G)$ is a monad and $\mathcal{S} = (G; F, \epsilon, G; \eta; F)$ is a comonad.*

PROOF: We shall prove that $\mathbb{T} = (F; G, \eta, F; \epsilon; G)$ is a monad. To this end note that $\eta_{G(F(A)}G(\epsilon_{F(A)} = 1_{G(F(A))}$ and $G(F(\eta_A))G(\epsilon_{F(A)}) = 1_{G(F(A)}$ by a straightforward application of the triangle equalities. The associativity condition translates to become:

$$\begin{tikzcd} G(F(G(F(G(F(X)))))) \arrow[r, "G(F(G(\epsilon_{F(X)})))"] \arrow[d, "G(\epsilon_{F(G(F(X)))})"] & G(F(G(F(X)))) \arrow[d, "G(\epsilon_{F(X)})"] \\ G(F(G(F(X)))) \arrow[r, "G(\epsilon_{F(X)})"'] & G(F(X)) \end{tikzcd}$$

which is naturality! $\square$

2.3.2 A presentation of monads due to Manes

There are a variety of ways of presenting monads. A rather important and economical way is due to Ernie Manes:

Proposition 2.3.2 *To give a monad on a category is to have an object map $T : \mathbb{X}_0 \longrightarrow \mathbb{X}_0$ a family of maps $\eta_X : X \longrightarrow T(X)$ and a "lifting" combinator*

$$\frac{X \xrightarrow{f} T(Y)}{T(X) \xrightarrow[\#(f)]{} T(Y)}$$

such that $\eta\#(f) = f$, $\#(\eta_X) = 1_{T(X)}$, and $\#(f)\#(g) = \#(f\#(g))$.

PROOF: Given a monad we define $\#(f) = T(f)\mu$ we must verify the equalities:

$$\begin{aligned}
\eta\#(f) &= \eta_X T(f)\mu_Y = f\eta_{T(Y)}\mu_Y = f \\
\#(\eta_X) &= T(\eta_X)\mu = 1_{T(X)} \\
\#(f)\#(g) &= T(f)\mu T(g)\mu = T(f)T(T(g))\mu\mu \\
&= T(f)T(T(g))T(\mu)\mu = T(fT(g)\mu)\mu = \#(f\#(g))
\end{aligned}$$

Conversely given this data we may define

$$\begin{aligned}
T(f) &= \#(f\eta) \\
\mu_X &= \#(1_{T(X)})
\end{aligned}$$

Note that $T(1_X) = \#(\eta_X) = 1_{T(X)}$ and $T(f)T(g) = \#(f\eta)\#(g\eta) = \#(f\eta\#(g\eta)) = \#(fg\eta) = T(fg)$. Thus, t is a functor. We now need to show η and μ are natural:

$$\begin{aligned}
\eta T(f) &= \eta\#(f\eta) = f\eta \\
T(T(f))\mu &= \#(\#(f\eta)\eta)\#(1) = \#(\#(f\eta)\eta\#(1)) = \#(\#(f\eta)) \\
&= \#(1)\#(f\eta) = \mu T(f)
\end{aligned}$$

Finally we need to verify the monad identities:

$$\begin{aligned}
\eta_{T(X)}\mu &= \eta\#(1) = 1_{T(X)} \\
T(\eta_X)\mu &= \#(\eta\eta)\#(1)\#(\eta\eta\#(1)) = \#(\eta) = 1_{T(X)} \\
T(\mu)\mu &= \#(\#(1)\eta)\#(1) = \#(\#(1)\eta\#(1)) \\
&= \#(\#(1)) = \#(1)\#(1) = \mu\mu
\end{aligned}$$

□

This is a considerable simplification of the monad laws. Despite its origin, this presentation has become known as the "Kleisli triple" presentation of a monad this is because this presentation is very closely tied to the composition in the Kleisli category. We shall shortly meet the Kleisli category of a monad. This presentation does allow us to outline some basic examples:

The list monad: Consider the list construction on sets: define $\eta(x) = [x]$ (the singleton list construction) and given $f : X \longrightarrow \mathrm{list}(Y)$ set

$$\#(f) : \mathrm{list}(X) \longrightarrow \mathrm{list}(Y); [x_1, ..., x_n] \mapsto f(x_1) +\!\!+ ... +\!\!+ f(x_n)$$

the Manes identities are then easy to check.

The bag monad: Every set may be used to generate sets of bags: a bag informally is set in which repetitions count. Thus, one way to present a bag is as a (finite) set of pairs $\{n_1 \cdot a_1, ..., n_r \cdot a_r\}$ where $n_i \in \mathbb{N}$ indicates the multiplicity of the element a_i. Thus, bags have a membership function which lands in $\mathbb{N}$ rather than bool. More formally a bag on X is just an element of the free commutative monoid generated by X. It is apparent, therefore, that $\mathcal{B}(X)$ is a monad. The unit is given by the singleton bag and the lifting of $f : X \longrightarrow \mathcal{B}(Y)$ is given by defining the membership function as follows:

$$y \in f^\sharp(\{n_1.x_1, ..., n_r \cdot x_r\}) = \sum_{i=1,..,r} n_i \cdot (y \in f(x_i))$$

The powerset monad: Every set may be included as singletons into the powerset on the set. Further given $f : X \longrightarrow \mathcal{P}(Y)$ we may define

$$\#(f) : \mathcal{P}(X) \longrightarrow \mathcal{P}(Y); S \mapsto \bigcup_{x \in S} f(x)$$

it is easy to check that this also is a monad. Clearly, if we restrict this monad to finite subsets we also obtain a monad $\mathcal{P}_f$.

The exception monad: By $E(X)$ we shall denote the set X with an extra point added. Given $f : X \longrightarrow E(Y)$ there is an obvious way of extenting the map to $\#(f) : E(X) \longrightarrow E(Y)$ by sending the added point of $E(X)$ to the added point of $E(Y)$. Again the required identities are easily checked.

The M-set monad: Given that M is a monoid in sets set $M(X) = M \times X$ and $\eta(x) = (e, x)$ given $f : X \longrightarrow M \times Y$ define $\#(f) : M \times X \longrightarrow M \times Y; (m_1, x) \mapsto (m_1 m_2, y)$ where $f(x) = (m_2, y)$. This is easily checked to be a monad.

Closure systems: Let P be a poset a monad on P amounts to a set map $J : P \longrightarrow P$ such that $x \leq J(x)$ and $x \leq J(y)$ implies $J(x) \leq J(y)$. Such a J is often called a **closure system**. Note that $J(x) = J(J(x))$ as $J(x) \leq J(J(X))$ and as $J(x) \leq J(x)$ also $J(J(x)) \leq J(x)$ and J is monotone.

Observe that the first three monads above (with the third being the finite powerset monad) are clearly related!

Given that monads and comonads are generated by adjoints a reasonable question concerns whether monads always arise through adjoints. For Set-enriched categories this is certainly true because we can construct adjunctions with this property. Below we shall give two such constructions which are extremal.

2.3.3 The Kleisli construction

Given a monad $\mathbb{T} = (T, \eta, \mu)$ on $\mathbb{X}$ the **Kleisli category**, $\mathbb{X}_{\mathbb{T}}$, is constructed as follows:

Objects $X \in \mathbb{X}$

Maps

$$\frac{X \xrightarrow{f} T(Y) \quad \in \mathbb{X}}{X \xrightarrow[f]{} Y \quad \in \mathbb{X}_{\mathbb{T}}}$$

Identity

$$\frac{X \xrightarrow{\eta} T(X) \quad \in \mathbb{X}}{X \xrightarrow[\eta]{} X \quad \in \mathbb{X}_{\mathbb{T}}}$$

Composition

$$\frac{\dfrac{X \xrightarrow{f} Y \quad Y \xrightarrow{g} Z \quad \in \mathbb{X}_{\mathbb{T}}}{X \xrightarrow{f} T(Y) \quad Y \xrightarrow{g} T(Z) \quad \in \mathbb{X}}}{\dfrac{X \xrightarrow{f\#(g)} T(Z) \quad \in \mathbb{X}}{X \xrightarrow{f\#(g)} Z \quad \in \mathbb{X}_{\mathbb{T}}}}$$

Proposition 2.3.3 $\mathbb{X}_{\mathbb{T}}$ *is a category, furthermore, there is an adjunction* $F_{\mathbb{T}} \dashv U_{\mathbb{T}} : \mathbb{X} \longrightarrow \mathbb{X}_{\mathbb{T}}$ *whose induced monad is* $\mathbb{T}$*.*

PROOF: We must show that $\mathbb{X}_{\mathbb{T}}$ is a category and for this it suffices to show that the composition is associative and has identities. Using the Kleisli presentation this is almost immediate. For example associativity of the composition amounts to the fact that $\#(\#(f)g)h = \#(f)\#(g)h$.

To verify that there is an adjunction we shall check the universal property. Define $U_{\mathbb{T}}(f) = \#(f)$ then

$$\begin{tikzcd}
X \arrow[r, "\eta"] \arrow[dr, "f"'] & U_{\mathbb{T}}(X) = T(X) \arrow[d, "\#(f)"] \\
& U_{\mathbb{T}}(Y) = T(Y)
\end{tikzcd}$$

commutes so all that is left is to check uniqueness of the map f (itself) as the solution. So suppose $g : X \longrightarrow T(Y)$ is an alternative with $\eta\#(g) = f$ then $f = \eta\#(g) = g$! $\square$

The reader should tame this proof! A good idea in this regard is to work through the proof without assuming the Kleisli presentation due to Manes.

This may see like a rather simple construction but it is also surprisingly powerful. This is witnessed by the extensive use of monads (and, in particular, the Kleisli construction) in functional programming languages (as demonstrated particularly in the programming language Haskell). To illustrate the power of the Kleisli construction considerthe following two examples:

The powerset monad: What is the Kleisli category of the powerset monad? A map in this Kleisli category is of the form $f : X \longrightarrow \mathcal{P}(Y)$ However such a map uniquely corresponds to a relation $R \subseteq X \times Y$ and it is easy to see that the Kleisli composition is exactly relational composition. Thus, the Kleisli category of the powerset monad is exacly the category of relations.

The exception monad: What is the Kleisli category of the exception monad? A map in this Kleisli category is a map $f : X \longrightarrow E(Y)$ such a map can be completely described by what it does to those elements of X which do not go to the "exceptional" point. Thus such maps correspond to partial maps from X to Y. it is easy to see that the Kleisli composition is partial map composition. This means that the Kleisli category of the exception monad is the partial map category.

The Kleisli category has an important property with respect to the adjunctions which generate the same monad:

Proposition 2.3.4 *If the adjunction $(\eta, \epsilon) : F \dashv U : \mathbb{X} \longrightarrow \mathbb{Y}$ generates the monad $\mathbb{T} = (T, \eta, \mu)$ then there is a unique full and faithful functor $V : \mathbb{X}_{\mathbb{T}} \longrightarrow \mathbb{Y}$ such that $F_{\mathbb{T}}; V = F$ and $U_{\mathbb{T}} = V; U$ so that $\eta_{\mathbb{T}} = \eta$ and $\mu_{\mathbb{T}} = \mu$.*

PROOF: Define $V(X) = F(X)$ and $V(f) = F(f)\epsilon$ where $f : X \longrightarrow T(Y)$. We must check that identities and composition is peserved: for identities note $V(\eta_X) = F(\eta_X)\epsilon_X = 1_{F(X)}$ for composition:

$$V(fT(g)\mu) = F(fT(g)\mu)\epsilon = F(fG(F(g)))F(G(\epsilon))\epsilon = F(f)F(G(F(g)))\epsilon\epsilon = F(f)\epsilon F(g)\epsilon$$

To show that V is full and faithful note that

$$\begin{array}{c} V(X) \rightarrow V(Y) \\ \hline\hline F(X) \rightarrow F(Y) \\ \hline\hline X \rightarrow G(F(Y)) \\ \hline\hline X \rightarrow T(Y) \end{array}$$

$\square$

In general, a Kleisli category will not inherit many properties from its parent category and this makes cases in which structure is inherited of special interest. Notably any Kleisli category does inherits coproducts (see next chapter) from its parent but it is hard to think of any other properties which are automatic!

2.3.4 The Eilenberg-Moore construction

The other category of importance which also generates the adjunction is the **Eilenberg-Moore category** of the monad. Its object are called $\mathbb{T}$-algebras and its maps are homomorphisms of these $\mathbb{T}$-algabras..

Given a monad $\mathbb{T}$ on $\mathbb{X}$ a **$\mathbb{T}$-algebra** (X, ν) is an object X together with a **structure map** $\nu : T(X) \longrightarrow X$ such that

$$\begin{tikzcd} X \arrow[r, "\eta"] \arrow[dr, equal] & T(X) \arrow[d, "\nu"] \\ & X \end{tikzcd} \qquad \begin{tikzcd} T(T(X)) \arrow[r, "\mu"] \arrow[d, "T(\nu)"'] & T(X) \arrow[d, "\nu"] \\ T(X) \arrow[r, "\nu"'] & X \end{tikzcd}$$

A **homomorphism** of T-algebras $f : (X, \nu) \longrightarrow (X', \nu')$ is a map $f : X \longrightarrow X'$ in $\mathbb{X}$ such that

$$\begin{tikzcd} T(X) \arrow[r, "T(f)"] \arrow[d, "\nu"'] & T(X') \arrow[d, "\nu'"] \\ X \arrow[r, "f"'] & X' \end{tikzcd}$$

Clearly these $\mathbb{T}$-algebras and their homomorphisms form a category, this category is is denoted $\mathbb{X}^{\mathbb{T}}$ and is the category of **Eilenberg-Moore algebras**:

Proposition 2.3.5 *The Eilenberg-Moore category of algebra $\mathbb{X}^{\mathbb{T}}$ has an underlying functor $U^{\mathbb{T}} : \mathbb{X}^{\mathbb{T}} \longrightarrow \mathbb{X}$ which has a left adjoint such that the induced monad is precisely $\mathbb{T}$.*

PROOF: The underlying functor has $U^{\mathbb{T}}(X, \nu) = X$ and $U^{\mathbb{T}}(f) = f$. Given any $X \in \mathbb{X}$ the pair $(T(X), \mu_X)$ is clearly a $\mathbb{T}$-algebra (it is called the **free** $\mathbb{T}$-algebra). We have:

$$\begin{tikzcd} X \arrow[r, "\eta"] \arrow[dr, "f"'] & U(T(X), \mu) \arrow[d, dotted, "U(T(f)\nu')"] \\ & (X', \nu') \end{tikzcd}$$

where note $\eta T(f)\nu' = f\eta\nu' = f$ so the diagram commutes. Furthermore, $T(f)\nu'$ is a morphism of algebras as

$$\begin{tikzcd} T(T(X)) \arrow[r, "\mu"] \arrow[d, "T(T(f))"'] & T(X) \arrow[d, "T(f)"] \\ T(T(X')) \arrow[r, "\mu"] \arrow[d, "T(\nu')"'] & T(X') \arrow[d, "\nu'"] \\ T(X') \arrow[r, "\nu'"'] & X' \end{tikzcd}$$

Thus, $T(f)\nu'$ is a a candidate for the universal map. Suppose h also works then $\eta h = f$ and:

$$\begin{tikzcd} T(T(X)) \arrow[r, "\mu"] \arrow[d, "T(h)"'] & T(X) \arrow[d, "h"] \\ T(X') \arrow[r, "\nu'"'] & X' \end{tikzcd}$$

Thus, $h = T(\eta)\mu h = T(\eta)T(h)\nu' = T(\eta h)\nu' = T(f)\nu'$.

Finally it is easy to check that the monad this induces is the orignal monad. $\square$

As for the Kleisli category the Eilenberg-Moore category enjoys a special property:

Proposition 2.3.6 *If the adjunction $(\eta, \epsilon) : F \dashv U : \mathbb{X} \longrightarrow \mathbb{Y}$ generates the monad $\mathbb{T} = (T, \eta, \mu)$ then there is a unique functor $W : \mathbb{Y} \longrightarrow \mathbb{X}^{\mathbb{T}}$ such that $W; U^{\mathbb{T}} = U$ and $F; W = F^{\mathbb{T}}$ so that $\eta^{\mathbb{T}} = \eta$ and $\mu^{\mathbb{T}} = \mu$.*

PROOF: Define $W(Y) = (U(F(U(Y))) \xrightarrow{U(\epsilon)} U(Y))$. We must check that this is an algebra which entails checking that

$$
\begin{tikzcd}
U(Y) \arrow[r, "\eta"] \arrow[dr, equal] & U(F(U(Y))) \arrow[d, "U(\epsilon)"] \\
& U(Y)
\end{tikzcd}
\qquad
\begin{tikzcd}
U(F(U(F(U(Y))))) \arrow[r, "\mu=U(\epsilon)"] \arrow[d, "U(F(U(\epsilon)))"] & U(F(U(Y))) \arrow[d, "U(\epsilon)"] \\
U(F(U(Y))) \arrow[r, "U(\epsilon)"'] & U(Y)
\end{tikzcd}
$$

commute. The first is a triangle identity the second is naturality.

Given $f : Y_1 \longrightarrow Y_2$ we have $U(f) : U(Y_1) \longrightarrow U(Y_2)$ and naturality supplies the fact that this induces an algebra homomorphism between $(U(Y_1), U(\epsilon_{Y_1}))$ and $(U(Y_1), U(\epsilon_{Y_1}))$. This is therefore a functor and the above properties are easily checked.

It remains to show that the functor is unique. So we suppose K' is another such functor. $K'(Y)$ must land on an object with underlying object $U(Y)$ so that it is forced to be of an algebra of the form $T(U(Y)) = U(F(U(Y))) \xrightarrow{\nu'} U(Y)$.

Now K' carries $F(U(Y))$ to an algebra $K'(F(U(Y)) = F^{\mathbb{T}}(U(Y))$ so that the algebra is actually $(T(U(Y)), \mu)$ and the algebra homomorphism $\epsilon : F(U(Y)) \to Y$ becomes a homomorphism:

$$
\begin{tikzcd}
T(T(U(Y))) \arrow[r, "T(U(\epsilon))"] \arrow[d, "\mu"'] & T(U(Y)) \arrow[d, "\nu'"] \\
T(U(Y)) \arrow[r, "U(\epsilon)"'] & U(Y)
\end{tikzcd}
$$

However preceding this square with $T(\eta)$ shows $\nu' = U(\epsilon)$ showing $K = K'$. $\square$

What structure does the Eilenberg-Moore category of algebras inherit from its parent category? The answer is a great deal. We shall delay a fuller discussion of this until we have discussed limits and colimits in the next section. However, we shall discuss one example here: factorizations. If the parent category has a factorization system (for example Set has its epic-monic factorization) and the functor T preserves this factorization – in fact simply preserves $\mathcal{E}$-maps – then the category of algebras inherits the factorization. In the case of Set this observation turns out to be quite important as it actually implies that *all* categories of algeras over sets inherit its epic-monic factorization. We start with:

Lemma 2.3.7 *If $\mathbb{X}$ has and $\mathcal{E}$-$\mathcal{M}$ factorization system and $\mathbb{T} = (T, \eta, \mu)$ is such that T preserves the factorization $\mathcal{E}$-maps (that is $T(e) \in \mathcal{E}$ whenever $e \in \mathcal{E}$) then $\mathbb{X}^{\mathbb{T}}$ inherits the $\mathcal{E}$-$\mathcal{M}$ factorization system.*

PROOF: A homomorphism of algebras $f : (X_1, \nu_1) \longrightarrow (X_2, \nu_2)$ is in $\mathcal{E}$ (respectively $\mathcal{M}$) in case f

is.We need to show that so defined these sets of maps are orthogonal in $\mathbb{X}^{\mathbb{T}}$.

$$
\begin{tikzcd}
& T(X_1) \arrow[rr, two heads, "T(e)"] \arrow[dl, "\nu_1"'] \arrow[dd, "T(f)"] & & T(X_2) \arrow[dl, "\nu_2"] \arrow[dd, "T(f')"] \arrow[ddll, dotted] \\
X_1 \arrow[rr, two heads, "e"] \arrow[dd, "f"'] & & X_2 \arrow[dd, "f'"] \arrow[ddll, dotted, "k"] & \\
& T(X_3) \arrow[rr, tail, "T(m)"] \arrow[dl, "\nu_3"'] & & T(X_4) \arrow[dl, "\nu_3"] \\
X_3 \arrow[rr, tail, "m"'] & & X_4 &
\end{tikzcd}
$$

In the above the cross-map k is *always* guaranteed to exist and be unique. The difficulty is to show that it is an algebra homomorphism: that is $T(k)\nu_3 = \nu_2 k$. However, note that both these are potential cross-maps for the square connecting $T(e)$ and m. As $T(e)$ is $\mathcal{E}$ this is unique and so the desired equality holds.

Next we need to show that every map can be factorized. This means that the image can be endowed with a $\mathbb{T}$-algebra structure map. The candidate is given by the obvious cross map:

$$
\begin{tikzcd}
T(X_1) \arrow[r, two heads, "T(e(f))"] \arrow[d, "\nu_1"'] & T(\mathrm{Im}(f)) \arrow[r, tail, "T(m(f))"] \arrow[d, dotted, "\nu"] & T(X_2) \arrow[d, "\nu_2"] \\
X_1 \arrow[r, "e(f)"'] & \mathrm{Im}(f) \arrow[r, tail, "m(f)"'] & X_2
\end{tikzcd}
$$

where we must show that the required diagrams for this to be an algebra homomorphism must hold. However, these are easily seen to follow from the orthogonality property. $\square$

The immediate consequence of this is that for Set *all* its categories of algebras inherit the epic monic factorization because epics in Set are always retracts and so necessarily are preserved by all functors.

Corollary 2.3.8 *All Eilenberg-Moore categories over sets have an epic-monic factorization system.*

Later we shall see this, in fact, is a *regular* epic-monic factorization. To see that this is a non-trivial observation it is worth considering the category of compact Haussdorf spaces: this is the Eilenberg-Moore category of algebras for the ultra-filter monad. This means that in this category the image of a continuous map is a compact Haussdorf subspace of the codomain so is, in particular, closed. This ensures that such homomorphisms are "proper" – these are rather sophisticated topological properties even if they are well-known in this situation.

2.3.5 Exercises

1. Prove carefully that the (finite) powerset monad is really a monad! Describe the Kleisli category for the (finite) powerset monad (hint: relations). What are the Eilenberg-Moore algebras for the (finite) powerset monad? (Hint: semi-lattices).

2. Prove carefully that the exception monad on sets is really a monad! Describe the Kleisli category (hint: partial maps). What are the Eilenberg-Moore algebras for the exception monad?

3. Prove carefully that the list monad on sets is a monad. Show that the Eilenberg-Moore category is exactly the category of monoids. Provide a description of the Kleisli category.

4. Prove carefully that the bag monad is really a monad! What are the Eilenberg-Moore algebras for the bag monad?

5. Prove that the state monad is a monad: on object it is defined as $\mathsf{St}(X) = (S \times X)^S$ where

$$\eta : X \longrightarrow (S \times X)^S; x \mapsto \lambda s.(s, x)$$

and

$$\mu : (S \times (S \times X)^S)^S \longrightarrow (S \times X)^S; f \mapsto \lambda s.(\lambda(s', f').f's')(fs)$$

Give an (alternative) description of the Kleisli category. (Harder) what is an Eilenberg-Moore algebra for the state monad?

6. (Hard) The **filter monad** on sets is defined as

$$\mathcal{F}(X) = \{U \subseteq \mathcal{P}(X) | X \in U, \forall u, v \in U. \Rightarrow u \cap v \in U, \forall u \in U, v \in \mathcal{P}(X).u \subseteq v \Rightarrow v \in U\}$$

that is $\mathcal{F}(X)$ is the set of **filters** in the powerset of X. A filter is a set of subsets which is upward closed, that is contains all supersets of its members, and contains the intersection of any finite set of its members. In particular this means a filter must contain the full set as it must contain the intersection of the empty set of its members (which is the full set). The set of all subsets – including the empty set – is clearly a filter (a filter is said to be *proper* if it is not this one). The unit of the monad is

$$\eta : X \longrightarrow \mathcal{F}(X); x \mapsto \{X' \subset X | x \in X'\}$$

takes an element x to the principal filter generated by $\{x\}$. Given a map $f : X \longrightarrow \mathcal{F}(Y)$ we may construct a map $f^\sharp : \mathcal{F}(X) \longrightarrow \mathcal{F}(Y)$ where $f^\sharp(U) = \{Y' \subseteq Y | f^{-1}(\Box Y') \in U\}$ where $\Box Y' = \{V \in \mathcal{F}(Y) | Y' \in V\}$.

Prove that this defines a monad.

(Harder) Prove that the algebras of this monad are precisely continuous lattices! A continuous lattice has all meets (infima) and has joins (suprema) of directed sets (recall a subset of a partially ordered set is *directed* in case it is nonempty and every pair of elements u and v from the set is dominated, that is there is a z in the set with $u \leq z$ and $v \leq z$). The morphisms must clearly preserve this structure. Given a continuous lattice X notice that there is a canonical structure map defined by:

$$\nu : \mathcal{F}(X) \longrightarrow X; U \mapsto \bigvee_{u \in U} \bigwedge_{x \in u} x$$

7. (Hard) The **ultra-filter monad** on sets is defined as

$$\mathcal{U}(X) = \{U \subseteq \mathcal{P}(X) | U \in \mathcal{F}, \forall X' \subseteq X. X' \in U \vee X \backslash X' \in U\}$$

that is $\mathcal{U}(X)$ is the set of ultra filters on X. An ultra-filter is a proper filter which for each subset contains either it or its complement. Clearly this implies that the filter is a maximal proper filter (and actually this characterizes maximal proper filters). As for the filter monad the unit on x picks out the principal filter containing $\{x\}$. The lifting map is defined in the same manner as for filters.

Prove this is a monad. (Harder) Prove that the algebras of this monad are compact Hausdorff spaces! Toward this end it is useful to realize that each compact Hausdorff space comes with with a canonical ultra-filter structure map as each ultra-filter converges on such a space to a unique point. Conversely, the convergence properties of such a space determine it.

8. Prove that for monads (T, η_T, μ_T) and (S, η_S, μ_S) and their Kleisli categories the following square of functors commute

$$\begin{tikzcd} \mathbf{X}_T \arrow[r, "J"] & \mathbf{Y}_S \\ \mathbf{X} \arrow[u, "F_T"] \arrow[r, "K"'] & \mathbf{Y} \arrow[u, "F_S"'] \end{tikzcd}$$

if and only if there is a "distributive law" that is a natural transformation

$$\alpha : T; K \longrightarrow K; S$$

such that

$$(\eta_T; K)\alpha = K; \eta_S \qquad (\mu_T; K)\alpha = (T; \alpha)(\alpha; S)(K; \mu_S).$$

For each of the following monads, the exception monad on sets, the list monad on sets, and the powerset monad on sets say whether:

(a) the product functor lifts to the Kleisli category (describe it),

(b) the list functor lifts to the Kleisli category (describe it).

Chapter 3

Limits and colimits

Perhaps the most basic and most pervasive examples of universal and couniversal pairs arise through limits and colimits. We take time now to discuss this source of universal property.

The first section of this chapter can be read in parallel with the previous chapter as it relies for the mostpart on having only a basic understanding of what a universal and couniversal pair is. The last two sections, however, rely on having an understanding of adjoints. For example, one of the main results of the last section is that all Eilenberg-Moore categories over the category of sets are exact, complete and cocomplete.

3.1 Basic limits and colimits

We start by discussing the properties of initial and final objects, and products and coproducts.

3.1.1 Initial and final objects

An **initial object** in a category $\mathbb{C}$ is an object which has exactly one map to every object (including itself) in the category. We shall often denote an initial object as the numeral 0 to remind us that it is a starting point and denote the unique map as $?_A : 0 \longrightarrow A$.

In Set the initial object is the empty set, in vector spaces it is the 0-dimensional vector space, and in Cat it is the empty category.

Dual to an initial object is a **final object**: a final object in a category $\mathbb{C}$ is an object to which every object has exactly one map. We shall often denote the final object by the numeral 1 and the unique map by $!_A : A \longrightarrow 1$.

In Set the final object is the one element set, in vector spaces the final object is the same as the initial object (that is the 0-dimensional vector space) and in Cat it is the category with one object and one arrow.

A simple observation is:

Lemma 3.1.1 *If K and K' are initial in $\mathbb{C}$ then there is a unique isomorphism $\alpha : K \longrightarrow K'$.*

Proof: As K is initial there is exactly one map $\alpha : K \longrightarrow K'$. Conversely, as K' is initial there is a unique map $\alpha' : K' \longrightarrow K$. This map is the inverse of α as $\alpha\alpha' : K \longrightarrow K$ is the unique endo-map on K namely the identity and similarly we obtain $\alpha'\alpha = 1'_K$. $\Box$

Thus initial objects (and by duality final objects) are unique up to unique isomorphism.

While there can only be one map to a final object there can be many maps from a final object to a given object (consider Set for example). These maps are often called **elements** and we make the following observation:

Lemma 3.1.2 *Elements in any category are sections.*

PROOF: An element is a map $a : 1 \longrightarrow A$ and has a right inverse $! : A \longrightarrow 1$. $\square$

One way to view an object in a category is as a functor from the final category.

Proposition 3.1.3 *An object K in a category $\mathbb{C}$ is an initial object if and only if $K \dashv ! : 1 \longrightarrow \mathbb{C}$.*

PROOF: Consider the universal diagram

$$\begin{tikzcd} \star \arrow[r, "1_\star"] \arrow[dr, "1_\star"'] & !(K) \arrow[d, dotted, "!(1_\star^\sharp)"] \\ & !(X) \end{tikzcd}$$

As the final category only has one map each object must have a unique map from the initial object.

It is also clear that the counit is an isomorphism as it is the only map of the final category. This means that K, the functor, is full and faithful. $\square$

3.1.2 Binary products and coproducts

Perhaps one of the most fundamental structures a category can have is a product. In fact, it is so fundamental we have already assumed several times that the reader knew what a product was in order to facilitate the development. Let A and B be any two objects in a category then a product of A and B is an object, often written $A \times B$ equipped with two maps $\pi_0 : A \times B \longrightarrow A$ and $\pi_1 : A \times B \longrightarrow B$ such that given any object W with two maps $f : W \longrightarrow A$ and $g : W \longrightarrow B$ there is a unique map, often written $\langle f, g\rangle : W \longrightarrow A \times B$, such that $\langle f, g\rangle\pi_0 = f$ and $\langle f, g\rangle\pi_1 = g$.

The maps π_0 and π_1 are called **projections**.

This can be depicted graphically as

$$\begin{tikzcd} & & A \\ W \arrow[rru, "f"] \arrow[rrd, "g"'] \arrow[r, dotted, "\langle f,g\rangle"] & A \times B \arrow[ru, "\pi_0"'] \arrow[rd, "\pi_1"] & \\ & & B \end{tikzcd}$$

The product in Set is the cartesian product, in vector spaces is what is often called the "direct sum" (the dimensions of the vector spaces are added), and in Cat is the product of categories.

This time we shall start by describing the couniversal property which characterizes the product from this we can then interpret the sense in which products are unique. To state the universal property we need to consider the diagonal functor $\Delta : \mathbb{C} \longrightarrow \mathbb{C} \times \mathbb{C} : f \longrightarrow (f, f)$ and the meaning of a couniversal object at (A, B) for Δ:

$$\begin{tikzcd}
\Delta(C) \arrow[d, dotted, "\Delta(\langle f,g \rangle)"'] \arrow[dr, "(f,g)"] & \\
\Delta(A \times B) \arrow[r, "(\pi_0,\pi_1)"'] & (A, B).
\end{tikzcd}$$

This gives us immediately that:

Lemma 3.1.4 *The product of A and B is the couniversal pair at (A, B) for the diagonal functor.*

The uniqueness of couniversal pairs tells us the sense in which a product is unique up to a unique isomorphism. It is worth unwinding this statement more explicitly to give a direct proof. The result very closely follows the proof style given for the initial object:

Suppose $(K, (a_0, a_1))$ and $(K', (a'_0, a'_1))$ are both products of A and B then there is a unique map $\alpha : K \longrightarrow K'$ such that $\alpha a'_0 = a_0$ and $\alpha a'_1 = a_1$ as K' is a product. But similarly there is a unique map $\alpha' : K' \longrightarrow K$ such that $\alpha' a_0 = a'_0$ and $\alpha' a_1 = a'_1$. The composite $\alpha\alpha' : K \longrightarrow K$ has $\alpha\alpha' a_0 = \alpha a'_0 = a_0$ and $\alpha\alpha' a_1 = \alpha a'_1 = a_1$. However $1_K : K \longrightarrow K$ also has this property so we must conclude $\alpha\alpha' = 1_K$ and similarly $\alpha'\alpha = 1_{K'}$.

A product has a number maps associated with it. The first map we consider is the diagonal map $\Delta = \langle 1_A, 1_A \rangle : A \longrightarrow A \times A$ this is the unit of the adjunction implied by the existence of all couniversal pairs.

Observe that $\langle 1_A, 1_A \rangle \pi_0 = 1_A$ this gives:

Lemma 3.1.5 *The diagonal map if it exists for an object in any category is a section and therefore monic.*

On might think that the projections must be epimorphisms: this is clearly so for cartesian powers ($A \times A$, $A \times A \times A$, ..) but is definitely not the case in general. Even in Set there is a counterexample to this: $A \times 0 \xrightarrow[\pi\ 0]{} A$ for nonempty A is not surjective and thus not epic.

Of course $\_ \times \_$ is a functor $f \times g$ is define to be $\langle \pi_0 f, \pi_1 g \rangle$ as illustrated by:

$$\begin{tikzcd}
& A \arrow[rr, "f"] & & A' \\
A \times B \arrow[ur, "\pi_0"] \arrow[dr, "\pi_1"] \arrow[rr, dotted, "f \times g"] & & A' \times B' \arrow[ur, "\pi_0"] \arrow[dr, "\pi_1"] & \\
& A' \arrow[rr, "g"'] & & B'
\end{tikzcd}$$

Next for any binary product there is a symmetry map:

$$\begin{tikzcd}
& & A \\
A \times B \arrow[urr, "\pi_0"] \arrow[drr, "\pi_1"'] \arrow[r, dotted, "c_{AB}"] & B \times A \arrow[ur, "\pi_1"] \arrow[dr, "\pi_0"] & \\
& & B
\end{tikzcd}$$

It is not hard to check that $c_{AB}c_{BA} = 1_{A\times B}$ and therefore we have:

Lemma 3.1.6 *The symmetry map for any product is an isomorphism.*

and $\Delta_A c_{AA} = \Delta_A$: the trick is to break the maps into their components by post-composing them with the projections.

Lastly there is an important map which allows one to re-associate products:

$$a_{ABC} : (A \times B) \times C \longrightarrow A \times (B \times C)$$

this map is the unique map determined by the equations

$$a_{ABC}\pi_0 = \pi_0\pi_0 \qquad a_{ABC}\pi_1\pi_0 = \pi_0\pi_1 \qquad a_{ABC}\pi_1\pi_1 = \pi_1.$$

This map has an obvious inverse:

$$a^{-1}_{ABC} : A \times (B \times C) \longrightarrow (A \times B) \times C$$

determined by the equations

$$a^{-1}_{ABC}\pi_0\pi_0 = \pi_0 \qquad a^{-1}_{ABC}\pi_0\pi_1 = \pi_0\pi_0 \qquad a^{-1}_{ABC}\pi_1 = \pi_1\pi_1.$$

Lemma 3.1.7 *The associativity map for any product is an isomorphism.*

In many structures which are "product like" these isomorphisms are present even though the projections or diagonals are absent. The "coherence" diagrams satisfied by these isomorphisms then becomes significant.

We shall say that a category $\mathbb{C}$ **has binary products** if every pair of objects has a product:

Proposition 3.1.8 *The following are equivalent:*

(i) *A category $\mathbb{C}$ has binary products;*

(ii) *The diagonal functor is a left adjoint:*

$$(\Delta, (\pi_0, \pi_1)) : \Delta \dashv \_ \times \_ : \mathbb{C} \longrightarrow \mathbb{C} \times \mathbb{C};$$

(iii) *There is and object operation $(A, B) \mapsto A \times B$ on $\mathbb{C}$ with two families of maps $\pi_0^{A,B} : A \times B \longrightarrow A$ and $\pi_1^{A,B} : A \times B \longrightarrow B$ together with a pairing combinator:*

$$\frac{f : X \longrightarrow A \qquad g : X \longrightarrow B}{\langle f, g\rangle : X \longrightarrow A \times B}$$

such that

$$\langle f, g\rangle\pi_0 = f \qquad \langle f, g\rangle\pi_1 = g \qquad \langle \pi_0, \pi_1\rangle = 1_{A\times B} \qquad h\langle f, g\rangle = \langle hf, hg\rangle.$$

PROOF: The first two formulations given the view of products as a couniversal pair are clearly equivalent. For the final formulation it is straightforward to check that a product satisfies these identities but less obvious that these define a product. To establish this we show that the pairing map is unique: suppose that k has $k\pi_0 = f$ and $k\pi_1 = g$ then

$$k = k1_{A\times B} = k\langle \pi_0, \pi_1\rangle = \langle k\pi_0, k\pi_1\rangle = \langle f, g\rangle.$$

□

The equality $\langle \pi_0, \pi_1\rangle = 1_{A\times B}$ is called to **surjective pairing** requirement. It is this identity which often is the most difficult to secure. For example, while it is possible in the lambda calculus to define many pairing combinators (satisfying everything but this last identity) it is provably impossible to produce a surjective pairing.

We should also note here that the formation of pairs can be reversed (which is the content of the surjective pairing condition, thus the inference rule may be written as a two-way inference:

$$\frac{h\pi_0 = f : X \longrightarrow A \qquad h\pi_1 = g : X \longrightarrow B}{\overline{h = \langle f, g\rangle : X \longrightarrow A \times B}}$$

Dual to the notion of a product is the notion of a coproduct. Let A and B be any two objects in a category then a **coproduct** of A and B is an object, often written $A + B$ equipped with two maps $\iota_0 : A \longrightarrow A+B$ and $\iota_1 : B \longrightarrow A+B$ such that given any object V with two maps $h : A \longrightarrow V$ and $k : W \longrightarrow B$ there is a unique map, often written $\langle h|k\rangle : W \longrightarrow A \times B$, such that $\iota_0\langle h|k\rangle = h$ and $\iota_1\langle h|k\rangle = k$.

The maps ι_0 and ι_1 are called the **coprojections**.

We express this universal property diagrammatically as follows:

$$\begin{tikzcd}
A \arrow[dr, "\iota_0"'] \arrow[drr, "h"] & & \\
& A+B \arrow[r, dotted, "\langle h|k\rangle"] & V \\
B \arrow[ur, "\iota_1"] \arrow[urr, "h"'] & &
\end{tikzcd}$$

Now it is interesting to note that the coproduct in Set is entirely different from the product. It is the disjoint union of the two sets the coprojections are the embeddings of the components into the disjoint union. In vector spaces the product and coproduct coincide — when this happens, in a nice way, we shall say that we have biproducts (see later). In Cat the coproduct is the disjoint union of the categories much as for sets.

The analogue to the diagonal maps is called the **codiagonal** and is the map

$$\nabla = \langle 1_A | 1_A\rangle : A + A \longrightarrow A.$$

We say a category **has coproducts** if every pair of objects has a coproduct. We therefore have:

Proposition 3.1.9 *The following are equivalent:*

(i) *A category $\mathbb{C}$ has binary coproducts;*

(ii) *The diagonal functor is a right adjoint:*

$$((\iota_0, \iota_1), \nabla) : \_ + \_ \dashv \Delta : \mathbb{C} \times \mathbb{C} \longrightarrow \mathbb{C};$$

(iii) *There is and object operation $(A, B) \mapsto A + B$ on $\mathbb{C}$ with two families of maps $\iota_0^{A,B} : A \longrightarrow A + B$ and $\iota_1^{A,B} : B \longrightarrow A + B$ together with a copairing combinator:*

$$\frac{f : A \longrightarrow X \qquad g : B \longrightarrow X}{\langle f|g\rangle : A + B \longrightarrow X}$$

such that

$$\iota_0\langle f|g\rangle = f \qquad \iota_1\langle f|g\rangle = g \qquad \langle \iota_0|\iota_1\rangle = 1_{A+B} \qquad \langle f|g\rangle h = \langle fh|gh\rangle.$$

This actually gives us a basic example of a functor which is both a left and right adjoint because we know Set has products and coproducts so $\Delta : \mathsf{Set} \longrightarrow \mathsf{Set} \times \mathsf{Set}$ must be both a left and right adjoint.

In the next chapter we shall discover that often the product and coproduct interact as they do in Set and obey a distributive law which states that the map

$$\langle 1 \times \iota_0 | 1 \times \iota_1\rangle : A \times B + A \times C \longrightarrow A \times (B + C)$$

which always exists, is an isomorphism. In general, however, it is not the case that we can expect such an interaction in general. In particular, notice that this makes for an arithmetic of objects which parallels the better known one for numbers. We would not therefore allow that $A \times B = A + B$ but, of course, this does happen for the product and coproduct in vector spaces.

3.1.3 Limits and colimits of diagrams

There is a general notion of the limit and colimit of a "diagram" which we now introduce. This leads to the notion of a complete category, that is a category in which the limits of all (small) diagrams exist. Of course, certain limits (e.g. equalizers and products) imply the presence of all other limits so that completeness can be reduced to having all equalizers and all (small) products.

A diagram is, concretely a morphism of a directed graph into a category. Occasionally this notion of a diagram is too restrictive as we may also want to say that certain composites in the diagram must be equal. A directed graph together with the specification that certain composites are equal is just a **presentation** of a category. We may construct an actual category from such a presentation by moving to the path category generated by the graph and "forcing" the equalities we wish to hold to be true. This we may do by generating the smallest congruence on the path category which includes the desired equalities. The category we want is then the quotient category with respect to this congruence and the diagrams are then functors from this category.

In this section we shall be ambivalent about whether we are working with functors from categories, presentations of categories, or just diagrams. The results we prove rely only on the directed graph structure and not on any overlying commuting constraints.

Thus, a **diagram** $D : \mathcal{G} \longrightarrow U(\mathbb{C})$ is a map of directed graphs into the underlying directed graph of $\mathbb{C}$ (we know this corresponds to a functor $D^\sharp : \mathsf{Path}(\mathcal{G}) \longrightarrow \mathbb{C}$) then a D–**cone** over this diagram consists of an object A, called the **apex** of the cone together with for each node N of $\mathcal{G}$ a map $\alpha_N : A \longrightarrow D(N)$ such that for each arrow of $\mathcal{G}$, $a : N_1 \longrightarrow N_2$, we have $\alpha_{N_1} G(a) = \alpha_{N_2}$.

A morphism of cones $(\alpha, h, \beta) : \alpha \longrightarrow \beta$ is given by a map in $\mathbb{C}$, $h : A \longrightarrow B$ between the apexes of the cones such that $\alpha_N = h\beta_N$ for all the nodes of the diagram. We observe:

Lemma 3.1.10 *The cones over $D : \mathcal{G} \longrightarrow \mathbb{C}$ form a category, $\mathsf{Cone}_D(\mathbb{C})$, with objects the cones and maps the morphisms of cones.*

This is straightforward to check and we leave it to the reader.

A **limit** of a diagram is a final object in the category $\mathsf{Cone}_D(\mathbb{C})$. We often write the apex of this cone as $\varprojlim(D)$ with **projections** $\pi_N : \varprojlim(D) \longrightarrow G(N)$.

We may display diagrammatically the concept of a limit of a diagram as follows:

$$\begin{tikzcd}
A \arrow[rrrr, dotted] \arrow[dr, "\alpha_{N_1}"'] \arrow[ddrr, "\alpha_{N_3}"] \arrow[drrr, "\alpha_{N_2}"] & & & & \lim(D) \arrow[dlll, "\pi_{N_1}"'] \arrow[dl, "\pi_{N_2}"] \arrow[ddll, "\pi_{N_3}"'] \\
& D(N_1) \arrow[rr, "D(a)"] & & D(N_2) & \\
& & D(N_3) \arrow[ur, "D(b)"'] & &
\end{tikzcd}$$

First we note that this notion of limit subsumes both the definition of a final object (take $\mathcal{G}$ to be the empty graph) and the definition of the product (take $\mathcal{G}$ to be the discrete graph – with no arrows – with two nodes). Furthermore it suggests what should be the definition of an n-ary product: namely, the limit of a discrete diagram with n nodes. However, as illustrated above there are many more shapes diagrams may take and, thus, a whole variety of limits of which we should develop some understanding.

Before doing so, however, it is useful to relate this notion of limit to that of a universal pair. To do this, as is often the case, it is a matter of choosing ones category and functors carefully.

Let $\mathsf{Dgrm}(\mathcal{G}, \mathbb{C})$ be the category of **$\mathcal{G}$-shaped diagrams** in $\mathbb{C}$ having objects morphisms of directed graphs $D : \mathcal{G} \longrightarrow U(\mathbb{C})$ and maps natural transformations (equivalently natural transformations between functors $D^\sharp : \mathsf{Path}(\mathcal{G}) \longrightarrow \mathbb{C}$). From general principles we know that this is a category.

There is also a diagonal functor $\Delta : \mathbb{C} \longrightarrow \mathsf{Dgrm}(\mathcal{G}, \mathbb{C})$ which takes an object C to the degenerate diagram on that object where all the nodes of $\mathcal{G}$ become the object C itself and the arrows of $\mathcal{G}$ become the identity map 1_C. We now have:

Proposition 3.1.11 $\mathsf{Dgrm}(\mathcal{G}, \mathbb{C})$ *is a category with a diagonal functor* $\Delta : \mathbb{C} \longrightarrow \mathsf{Dgrm}(\mathcal{G}, \mathbb{C})$. *Furthermore a couniversal pair at diagram* D *for* Δ *is exactly a limit of* D, $(\varprojlim D, \pi)$.

Proof: This is more a matter of translation than of proof! The first thing to realize is that a natural transformation $h : \Delta(A) \longrightarrow D$ is exactly a cone. Thus, the couniversal diagram:

$$\begin{tikzcd}
\Delta(A) \arrow[d, dotted, "\Delta(\langle f \rangle)"'] \arrow[dr, "f"] & \\
\Delta(\varprojlim(D)) \arrow[r, "\pi"'] & D.
\end{tikzcd}$$

also ensures that $(\varprojlim(D), \pi)$ is final in the category of cones. □

We shall say that a category is **complete** in case $\Delta : \mathbb{C} \longrightarrow \mathsf{Dgrm}(\mathcal{G}, \mathbb{C})$ has a right adjoint for each graph $\mathcal{G}$. Here the size of the directed graphs is important: we are tacitly assuming that the directed graphs we are talking about are "small" in the sense that both the nodes and arrows form sets. However, we could assume that both the node and arrow sets must be finite: we will say the category is **finitely complete** if for any finite $\mathcal{G}$ the functor Δ has a right adjoint.

Dual to the notion of limit is that of colimit. A colimit of a diagram D is a universal pair, often written $(\varinjlim D, \iota)$, at D for the functor Δ.

We may display a colimit diagrammatically as:

$$\begin{tikzcd}
& D(N_1) \arrow[rr, "D(a)"] \arrow[ddl, "\iota_{N_1}"'] \arrow[ddrrr, "\beta_{N_1}"] & & D(N_2) \arrow[ddlll, "\iota_{N_2}"] \arrow[ddr, "\beta_{N_2}"] & \\
& & D(N_3) \arrow[ul, "D(b)"'] \arrow[dll, "\iota_{N_3}"] \arrow[drr, "\beta_{N_3}"] & & \\
\lim D \arrow[rrrr, dotted] & & & & A
\end{tikzcd}$$

We shall say that a category is **cocomplete** in case $\Delta : \mathbb{C} \longrightarrow \mathsf{Dgrm}(\mathcal{G}, \mathbb{C})$ has a left adjoint for each small graph $\mathcal{G}$. We will say the category is **finitely cocomplete** if for any finite $\mathcal{G}$ the functor Δ has a left adjoint.

A useful fact is:

Lemma 3.1.12 *Left adjoints preserve colimits and right adjoints preserve limits.*

Proof: A colimit is a universal pair: universal pairs are preserved by left adjoints. □

3.1.4 Special limits: equalizers and pullbacks

There are many shapes for diagrams and in this section we start by focusing on two special shapes: equalizer diagrams and pullback diagrams. We show how the existence of limits for various shapes of diagrams imply the existence of limits for others, culminating in proving that all limits exist when products and equalizers are present.

An **equalizer diagram** is a parallel pair of arrows:

$$A \underset{g}{\overset{f}{\rightrightarrows}} B$$

a cone for the above equalizer diagram is called **an equalizer** of f and g and is given by an object Q together with a map $q : Q \longrightarrow A$ such that $qf = qg$ (this map is the map $Q \longrightarrow B$ required to make a cone). A limit (E, e) is called **the equalizer** even though it is not unique and satisfies the following couniversal property

$$\begin{tikzcd} Q \arrow[dr, dotted, "k"'] \arrow[drr, "q"] & & & \\ & E \arrow[r, "e"'] & A \arrow[r, shift left, "f"] \arrow[r, shift right, "g"'] & B \end{tikzcd}$$

that there is a unique k such that $ke = q$.

We shall often write the equalizer of f and g as the object couniversal pair $\pi_{f=g} : f = g \longrightarrow A$. We observe:

Lemma 3.1.13 *Suppose (E, e) is the equalizer of $A \underset{g}{\overset{f}{\rightrightarrows}} B$ then e is monic.*

PROOF: Suppose $xe = xy = z : X \longrightarrow A$ then $zf = zg$ so there is a unique $k : X \longrightarrow E$ such that $ke = z$. Therefore $k = x = y$ and so certainly $x = y$. □

A **pullback diagram** is a binary fan of arrows:

$$\begin{tikzcd} & A \arrow[d, "f"] \\ B \arrow[r, "g"'] & C \end{tikzcd}$$

a cone for the above pullback diagram is called **a pullback** of f and g and is given by an object Q together with two maps $q_A : Q \longrightarrow A$ and $q_B : Q \longrightarrow B$ such that $q_A f = q_B g$ (this map is the map $Q \longrightarrow C$ required to make a cone). A limit (E, e_A, e_B) is called **the pullback** even though it is not unique and satisfies the following couniversal property

$$\begin{tikzcd} Q \arrow[dr, dotted, "k"] \arrow[drr, "q_A"] \arrow[ddr, "q_B"'] & & \\ & E \arrow[r, "e_A"'] \arrow[d, "e_B"] & A \arrow[d, "f"] \\ & B \arrow[r, "g"'] & C \end{tikzcd}$$

that there is a unique k such that $ke_A = q_A$ and $ke_B = q_B$.

We shall occasionally write the pullback of f and g as the couniversal pair $(f \wedge g, \pi)$. Sometimes we shall want to view the pullback diagram in an asymmetric way and so we shall say (in the above) e_A is **the pullback along** f of g.

Lemma 3.1.14 *In any category*

(i) *the pullback of a monic along any map is a monic;*

(ii) *the pullback of a retraction along any map is a retraction;*

(iii) *the pullback of an isomorphism along any map is an isomorphism.*

PROOF:

(i) Suppose g is monic and $k_1 e_A = k_2 e_A$ then

$$k_1 e_B g = k_1 e_A f = k_2 e_A f = k_2 e_B g$$

so as $g$is monic $k_1 e_B = k_2 e_B$.

$$\begin{tikzcd}
Q \arrow[dr, dotted] \arrow[drr, "k_1e_A=k_2e_A"] \arrow[ddr, "k_1e_B=k_2e_b"'] & & \\
& E \arrow[r, "e_A"'] \arrow[d, "e_B"] & A \arrow[d, "f"] \\
& B \arrow[r, tail, "g"'] & C
\end{tikzcd}$$

However, this makes k_1 and k_2 comparison maps from the outer square to the pullback.

(ii) Suppose now g is a retraction so there is a g' with $g'g = 1_C$. Now consider the comparison map given by:

$$\begin{tikzcd}
A \arrow[dr, dotted, "k"] \arrow[drr, equal] \arrow[ddr, "fg'"'] & & \\
& E \arrow[r, "e_A"'] \arrow[d, "e_B"] & A \arrow[d, "f"] \\
& B \arrow[r, two heads, "g"'] & C
\end{tikzcd}$$

Clearly it is a left inverse for e_A.

(iii) As an isomorphism is a monic retraction we may combine the two previous parts to obtain this one.

□

A helpful observation is the following:

Lemma 3.1.15 *In any category $f : A \longrightarrow B$ is monic if and only if the following square is a pullback:*

$$\begin{tikzcd} A \arrow[r, equal] \arrow[d, equal] & A \arrow[d, "f"] \\ A \arrow[r, "f"'] & B. \end{tikzcd}$$

Proof: If this square is a pullback then whenever $xf = yf$ there is a unique comparison map

$$\begin{tikzcd} X \arrow[dr, dotted] \arrow[drr, "x"] \arrow[ddr, "y"'] & & \\ & A \arrow[r, equal] \arrow[d, equal] & A \arrow[d, "f"] \\ & A \arrow[r, "f"'] & B. \end{tikzcd}$$

which shows $x = y$. Conversely if f is monic then whenever we form the outer square so that it commutes $x = y$ so that this also gives a comparison map, whose uniqueness is forced by the fact that f is monic. $\square$

This then allows us to link the property of being monic to a couniversal property whence we can conclude:

Corollary 3.1.16 *Right adjoints preserve monics and (dually) left adjoints preserve epics.*

Pullback squares can be pasted together and the following observation is key to the behavior of squares under pasting with respect to pullbacks:

Proposition 3.1.17 *In the following (commuting) diagram:*

$$\begin{tikzcd} A \arrow[r, "f"] \arrow[d, "a"'] & B \arrow[r, "g"] \arrow[d, "b"'] & C \arrow[d, "c"] \\ A' \arrow[r, "f'"'] & B' \arrow[r, "g'"'] & C' \end{tikzcd}$$

(i) *if the two inner squares are pullbacks the outer square is a pullback;*

(ii) *if the rightmost square and outer square is a pullback the leftmost square is a pullback.*

Proof:

(i) Suppose $xc = yf'g' : X \longrightarrow C$ then there is a unique map $w : X \longrightarrow B$ such that $wg = x$ and $wb = yf'$. The latter equality gives a unique $v : X \longrightarrow A$ such that $va = x$ and $vf = w$. But this means $vfg = wg = x$. Thus, v is the desired comparison map. It remains to check uniqueness which is straightforward.

(ii) Suppose $xb = yf'$ Then $xgc = yf'g'$ so there is a unique k such that $ka = y$ and $kfg = xg$. But now both kf and x have the property that $xg = kfg$ and $xb = yf' = kaf' = kfb$. Thus, they both give the unique comparison map to the rightmost square, so $kf = x$. This means that k is a comparison map to the leftmost square: again it is straightforward to check that k is unique.

□

3.1.5 Completeness and cocompleteness

We shall say a category **has equalizers** or **has pullbacks** in case the appropriate limits always exist. The category of Set has both equalizers and pullbacks. Equalizers can be described very simply in Set, the limit cone for parallel set maps f and g is:

$$\{a \in A | f(a) = f(b)\} \subseteq A \underset{g}{\overset{f}{\rightrightarrows}} B.$$

Finite dimensional vector spaces over a field K (take, for example, $K = \mathbb{R}$) also have equalizers. These, however, this time are harder to describe but is basically the solution space for a set of linear equations. The following result shows that these examples also have pullbacks:

Proposition 3.1.18

(i) *If a category has binary products and equalizers then it has pullbacks;*

(ii) *If a category has pullbacks and binary products then it has equalizers;*

(iii) *If a category has a final object and pullbacks then it has binary products;*

(iv) *If a category has binary products then it has n-ary products for all $n > 0$.*

PROOF:

(i) We may construct the pullback as the equalizer:

$$f \wedge g \xrightarrow{\langle \pi_A, \pi_B \rangle} A \times B \underset{\pi_1 g}{\overset{\pi_0 f}{\rightrightarrows}} C$$

Where we notice that $\langle x, y \rangle \pi_0 f = \langle x, y \rangle \pi_1 g : X \longrightarrow C$ if and only if $xf = yg$ and the unique $k : X \longrightarrow f \wedge g$ has $k\langle \pi_A, \pi_B \rangle = \langle k\pi_A, k\pi_B \rangle = \langle f, g \rangle$ which is the case if and only if $k\pi_A = f$ and $k\pi_B = g$ as required.

(ii) The equalizer of two maps may be constructed using the following pullback:

$$\begin{tikzcd} f = g \arrow[r, "\pi"] \arrow[d, "\pi f"'] & A \arrow[d, "\langle f,g \rangle"] \\ B \arrow[r, "\Delta"'] & B \times B. \end{tikzcd}$$

Suppose $xf = xg : X \longrightarrow B$ then

$$xf\Delta = xf\langle 1, 1\rangle = \langle xf, xf\rangle = \langle xf, xg\rangle = x\langle f, g\rangle$$

so that there is a map $k : X \longrightarrow f = g$ with $k\pi = x$ which is forced to be unique as π is monic being a pullback of the section Δ.

(iii) The binary fan

$$\begin{tikzcd} & A \arrow[d, "!"] \\ B \arrow[r, "!"'] & 1 \end{tikzcd}$$

has the pullback the product of A and B as any cone of the one is a cone of the other.

(iv) The n-ary product may be built from a composite of binary products.

□

The main observation of this subsection is as follows:

Proposition 3.1.19 *A category is (finitely) complete if and only if it has (finite) products and equalizers.*

PROOF: If the category is (finitely) complete then it will have all (finite) products and equalizers so the content of the result is in the converse. Let $D : \mathcal{G} \longrightarrow U(\mathbb{C})$ be a diagram then we have the arrows of the graph G_1 and the nodes of the graph G_0 and we can form the following equalizer:

$$V \xrightarrow{\langle v_a\rangle_{a\in G_0}} \prod_{a\in G_0} D(a) \underset{\langle \pi_{\partial_1(f)}\rangle_{f\in G_1}}{\overset{\langle \pi_{\partial_0(f)}D(f)\rangle_{f\in G_1}}{\rightrightarrows}} \prod_{f\in G_1} D(\partial_1(f))$$

Notice that any equalizer of this equalizer is also a cone for D as

$$\begin{aligned} v_aD(f) &= \langle v_a\rangle_{a\in G_0}\pi_{\partial_0(f)}D(f)f \\ &= \langle v_a\rangle_{a\in G_0}\langle \pi_{\partial_0(g)}D(g)\rangle_{g\in G_1}\pi_f \\ &= \langle v_a\rangle_{a\in G_0}\langle \pi_{\partial_1(g)}\rangle_{g\in G_1}\pi_f \\ &= \langle v_a\rangle_{a\in G_0}\pi_{\partial_1(f)} \\ &= v_b. \end{aligned}$$

and conversely very cone for D gives rise to an equalizer of this diagram as

$$\begin{aligned} &\langle v_a\rangle_{a\in G_0}\langle \pi_{\partial_0(f)}D(f)\rangle_{f\in G_1} \\ &\quad= \langle\langle v_a\rangle_{a\in G_0}\pi_{\partial_0(f)}D(f)\rangle_{f\in G_1} \\ &\quad= \langle\langle v_{\partial_0(f)}D(f)\rangle_{f\in G_1} \\ &\quad= \langle\langle v_{\partial_1(f)}\rangle_{f\in G_1} \\ &\quad= \langle v_a\rangle_{a\in G_0}\langle \pi_{\partial_1(f)}\rangle_{f\in G_1}. \end{aligned}$$

Finally a morphism of cones over D becomes a morphism of equalizers as a map to a product is determined by its components. This means that the two limits have isomorphic cone categories. Thus if either has a limit it is also a limit of the other. □

3.1.6 Exercises

(1) Prove carefully that the symmetry and associativity maps for binary products are isomorphisms.

(2) If a category has products and coproducts prove that

$$\langle\langle f|g\rangle, \langle h|k\rangle\rangle = \langle\langle f,h\rangle|\langle g,k\rangle\rangle : A+B \longrightarrow C \times D.$$

(3) If a category $\mathbb{C}$ has products (and therefore a final object 1) prove that:

$$\begin{tikzcd} & A \arrow[dl, equal] \arrow[dr, "!"] & \\ A & & 1 \end{tikzcd}$$

is a limit cone. Show, therefore that $A \times 1$ and $1 \times A$ are canonically (in the sense that these maps always exist) isomorphic to A.

(4) If a category $\mathbb{C}$ has products prove that the following diagrams are always pullbacks:

$$\begin{tikzcd} A \times B \arrow[r, "f \times 1_B"] \arrow[d, "\pi_0"'] & A' \times B \arrow[d, "\pi_0"] \\ A \arrow[r, "f"'] & A' \end{tikzcd} \qquad \begin{tikzcd} A \arrow[r, "\langle 1_A, f\rangle"] \arrow[d, "f"'] & A \times B \arrow[d, "f \times 1_B"] \\ B \arrow[r, "\Delta"'] & B \times B. \end{tikzcd}$$

(5) If a category has products prove that

$$A \xrightarrow{\Delta} A \times A \underset{\pi_1}{\overset{\pi_0}{\rightrightarrows}} A$$

gives the equalizer of π_0 and π_1.

(6) Prove that in any category the pushout along any map of a section is itself a section. Show that the pullback of a section is not necessarily a section: give a counter example in finite sets.

(7) Prove that if $mr = 1_Y$ and $e = rm$ then, in the following diagram

$$Y \xrightarrow{m} X \underset{1_X}{\overset{e}{\rightrightarrows}} X \xrightarrow{r} Y$$

m is the equalizer of e and 1_X and r the coequalizer.

(8) Prove that if a functor $F : \mathbb{C} \longrightarrow \mathbb{D}$ preserves pullbacks and $\mathbb{C}$ has products that F preserves equalizers (Hint: show that $\langle F(p_0), F(p_1)\rangle$ is monic).

(9) Describe coequalizers in Set. Prove that Set is complete and cocomplete.

(10) Suppose that $(\eta, \epsilon) : F \dashv G : \mathbb{X} \longrightarrow \mathbb{Y}$ is a reflection (ϵ is a natural isomorphism) and that $\mathbb{X}$ has colimits then $\mathbb{Y}$ has colimits.

(11) Certain limits can be obtained from other limits:

(a) Prove that any category has limits for acyclic graphs with a source. These are graphs with an object from which any other object can be reached (in a unique way) following the direction of the arrows (hint: a tree in fact!).

(b) (Harder) Prove that any category with equalizers has limits for finite graphs with a source (any other object can be reached from the source in a not necessarily unique way following the direction of the arrow).

(c) Prove that if a category has pullbacks it has limits for all finite connected acyclic graphs. These are the diagrams which regarding each (directional) arrow as a two-way arrow leaves every object reachable from every other in a unique way.

(d) (Harder) Prove that if a category has pullbacks and equalizers that every finite connected diagram has a limit. These are the diagrams which regarding each (directional) arrow as a two way arrow leaves every object reachable from every other (possibly in many ways).

(12) Show that a small category (both objects and arrows are sets) which is complete is a preorder. Give a concrete proof that any meet preserving map from a complete poset has left adjoint.

(13) A complete partial order is a partial order with all limits or, equivalently all meets. Prove that the inclusion of complete partial orders into all partial orders is a reflection (you may need to look up some lattice theory here!).

(14) Show that $\mathsf{Set}^{\mathbb{C}^{\mathrm{op}}}$ for any small category $\mathbb{C}$ is complete and cocomplete.

(15) Show that a congruence on a 2-category is an equivalence relation on the 2-cells which only relates like typed 2-cells and satisfies:

(a) if $\alpha \sim \beta$ then $\gamma\alpha\gamma' \sim \gamma\beta\gamma'$ whenever these vertical composites are defined,

(b) if $\alpha \sim \beta$ then $\gamma; \alpha; \gamma' \sim \gamma; \beta; \gamma'$ whenever these horizontal composites are defined.

Show that every 2-functor can be factorized into a 2-functor which is surjective on 2-cells and bijective on 1-cells and 0-cells followed by a 2-faithful 2-functor (i.e one which is injective on the 2-cells hom-sets).

3.2 Limits, colimits and factorizations

When a category has both finite limits (or finite colimits) and a factorization system these two structures interact in a number of important ways. Regular categories, with their factorization of maps into regular epics monics, show how some very set-like properties can be captured in a purely categorical manner.

The section ends with a discussion of $\mathcal{M}$-extremal maps and reflexive factorizations. In particular the close correspondence between wide reflexive subcategories and reflexive factorizations is described.

3.2.1 Orthogonality and limits

Recall that given a class $\mathbb{A}$ of maps we may form the class $\mathbb{A}_\perp$ (or $\mathbb{A}_{\perp_w}$) of maps which are (weakly) right orthogonal to all the maps in $\mathbb{A}$. The following results may all be expressed in their dual form:

Lemma 3.2.1 *In any category with finite limits: the class of maps (weakly) right orthogonal to $\mathbb{A}$, $\mathbb{A}_\perp$ (respectively $\mathbb{A}_{\perp_w}$) have the following properties:*

(i) *The pullback of an $\mathbb{A}_\perp$–map (respectively $\mathbb{A}_{\perp_w}$–map) along any map is an $\mathbb{A}_\perp$–map (respectively $\mathbb{A}_{\perp_w}$–map);*

(ii) *If g_1 and g_2 are $\mathbb{A}_\perp$–maps (respectively $\mathbb{A}_{\perp_w}$–maps) then $g_1 \times g_2$ is an $\mathbb{A}_\perp$–map (respectively an $\mathbb{A}_{\perp_w}$–map);*

(iii) *Every section is in $\mathbb{A}_\perp$ (respectively $\mathbb{A}_{\perp_w}$) if and only if every the equalizing map $\pi_{f=g}$ for any parallel maps f and g is an $\mathbb{A}_\perp$–map (respectively $\mathbb{A}_{\perp_w}$–map).*

PROOF:

(i) Suppose we form the pullback of an $A_\perp$ (respectively $\mathbb{A}_{\perp_w}$) map g along a map h, then we must check that g', as below, is in $A_\perp$. To do this we consider a square comparing $f \in \mathbb{A}$ with g':

$$\begin{tikzcd}
X \arrow[r, "k_1"] \arrow[d, "f"'] & g \wedge h \arrow[r, "h'"] \arrow[d, "g'"] & A \arrow[d, "g"] \\
Y \arrow[r, "k_2"'] \arrow[ur, dotted, "v_2"] \arrow[urr, dotted, "v_1"] & B \arrow[r, "h"'] & C
\end{tikzcd}$$

Because g is (weakly) orthogonal we can conclude there is a map v_1 as shown. Next we can use v_1 to provide a unique map v_2 which is the desired cross map.

(ii) The fact that $g_1 \times g_2$ has the desired orthogonality property may be lifted straight from the fact that g_1 and g_2 have this property..

(iii) Such an equalizer can be calculated as the pullback of the diagonal map which is a section and therefore an $A_\perp$-map: by *(i)* it is therefore an $\mathbb{A}^\perp$-map.

Conversely a section is the equalizer of its idempotent and the identity.

□

A useful observation for orthogonal classes of maps across adjoints is as follows:

Lemma 3.2.2 *If $F \dashv G : \mathbb{X} \longrightarrow \mathbb{Y}$ and $\mathbb{A}$ is a class of maps in $\mathbb{X}$ while $\mathbb{B}$ is a class of maps in $\mathbb{Y}$ such that F preserves these chosen maps, that is $a \in \mathbb{A}$ implies $f(a) \in \mathbb{B}$, then G preserves the (weakly) right orthogonal maps in the sense that whenever $x \in \mathbb{B}_\perp$ (respectively $x \in \mathbb{B}_{\perp_w}$) then $g(x) \in \mathbb{A}_\perp$ (respectively $G(x) \in \mathbb{A}_{\perp_w}$).*

PROOF: Suppose $x \in \mathbb{B}_\perp$ then certainly testing against any $F(a)$ must produce the desired cross map as $F(a) \in \mathbb{B}$. However, we have the correspondence:

$$\frac{F(a) \longrightarrow x}{\overline{a \longrightarrow G(x)}}$$

where we are working in the arrow category so the top and bottom lines are commutative squares. This easily shows that $G(x)$ must be in $\mathbb{A}_\perp$. $\square$

3.2.2 Regular monics and extremal epics

A map is called a **regular monic** in case it is the equalizer of some pair of maps. Notice that if the category has colimits then for any map $f : A \longrightarrow B$ there is a universal pair of maps which that map equalizes. This pair (k_1, k_2), the **cokernel pair** of f, is given by pushing out the map along itself:

$$\begin{tikzcd} & B \arrow[dr, "k_1"] & \\ A \arrow[ur, "f"] \arrow[dr, "f"'] & & f \vee f \\ & B \arrow[ur, "k_2"'] & \end{tikzcd}$$

Thus a regular monic in a category with cokernel pairs is exactly a map which equalizes its own cokernel pair.

Now if every regular monic is an $\mathbb{A}^\perp$-map then we may conclude that every $\mathbb{A}$-map must be epic. For suppose $gh_1 = gh_2$ and $g \in \mathbb{A}$ then letting the equalizer of h_1 and h_2 be $k : K \longrightarrow B$ we have $g = g'k$ but as $k \in \mathbb{A}^\perp$ there is a unique map v unduced by othogonality:

$$\begin{tikzcd} A \arrow[r, "g"] \arrow[d, "g'"'] & B \arrow[d, equal] \arrow[dl, "v"'] \\ K \arrow[r, "k"'] & B \end{tikzcd}$$

which is necessarily an isomorphism as k is monic (because $kvk = k$ so $kv = 1_K$). Thus k is an isomorpism so that $h_1 = h_2$.

Lemma 3.2.3 $\mathbb{A}^\perp$ *contains all the regular monics if and only if all maps* $\mathbb{A}$ *are epic.*

PROOF: It remains to show that if all maps in $\mathbb{A}$ are epic that $\mathbb{A}^\perp$ contains all the regular monics. For this it suffices to show that every regular monic is orthogonal to every epic. Suppose m is regular monic and equalizes f and g, while e is epic, and suppose $eh_2 = h_1 m$:

$$\begin{tikzcd} A \arrow[r, "e"] \arrow[d, "h_1"'] & B \arrow[d, "h_2"] \arrow[dl, dotted, "k"] & \\ C \arrow[r, "m"'] & D \arrow[r, shift left, "f"] \arrow[r, shift right, "g"'] & E \end{tikzcd}$$

then $h_2f = h_2g$ as $eh_2f = h_1mf = h_1mg = eh_2g$ so that there is a unique map k as shown to the equalizer of f and g. The fact that it is the comparison map to the equalizer makes the upper triangle commute. For the lower triangle notice that $ekm = eh_2 = h_1m$ and m is monic. □

The last part of lemma 3.2.1 may be restated as saying $A_\perp$ contains all regular monics iff and only if $A_\perp$ contains all the sections. We may now use lemma 1.4.6 *(iii)* and the above observations to provide the following strengthening of that result:

Corollary 3.2.4 *The following are equivalent for a factorization system on a category with equalizers:*

(i) *The $\mathcal{M}$-maps are left-factor closed;*

(ii) *Every section is an $\mathcal{M}$-map;*

(iii) *Every regular monic is an $\mathcal{M}$-map;*

(iv) *Every $\mathbb{E}$-map is epic.*

We shall say that a factorization system is **stable** in case the pullback of a factorization is again a factorization. That is whenever the two squares are pullbacks $e'_f \in \mathbb{E}$ and $m'_f \in \mathcal{M}$

$$\begin{tikzcd}
f \wedge g \arrow[r, dotted, "e_{f'}"] \arrow[d, dotted, "g''"] & m_f \wedge g \arrow[r, "m_{f'}"] \arrow[d, dotted, "g'"] & X \arrow[d, "g"] \\
A \arrow[r, "e_f"'] & E_f \arrow[r, "m_f"'] & B
\end{tikzcd}$$

in view of lemma 3.2.1 it follows that it suffices to require that pullbacks of $\mathbb{E}$-maps along any map are $\mathbb{E}$-maps.

Lemma 3.2.5 *A factorization system on a category with pullbacks is stable if and only if pullbacks of $\mathbb{E}$-maps along any map are $\mathbb{E}$-maps.*

The standard "epi-mono" factorization of Set is stable.

Remark 3.2.6 If a factorization system is stable then if is easy to see that the products functor will preserve factorizations: that is whenever $f, g \in \mathcal{E}$ then $f \times g \in \mathcal{E}$. Notice that we necessarily have $f, g \in \mathcal{M}$ then $f \times g \in \mathcal{M}$ for every factorization system by the above result. We observe that:

Suppose that $\mathbb{X}$ has products and a factorization which is preserved by the products. Consider an enriched functor $F : \mathbb{C} \to \mathbb{D}$ between two $\mathbb{X}$-enriched categories then we may factorize the functor and obtain an intermediate enriched category as follows:

$$\begin{tikzcd}
\mathbb{C}(A,B) \times \mathbb{C}(B,C) \arrow[r, "m_{ABC}"] \arrow[d, "m(f_{AB}) \times m(f_{BC})"'] & \mathbb{C}(A,C) \arrow[d, "m(f_{AC})"'] \\
I_{f_{AB}} \times I_{f_{BC}} \arrow[r, dotted, "m"] \arrow[d, "e(f_{AB}) \times e(f_{BC})"'] & I_{f_{BC}} \arrow[d, "e(f_{AC})"'] \\
\mathbb{D}(F(A),F(B)) \times \mathbb{D}(F(B),F(C)) \arrow[r, "m"'] & \mathbb{D}(F(A),F(C))
\end{tikzcd}$$

This then gives a factorization on enriched categories which, furthermore, itself is preserved by products (as the product of an enriched category is formed using the product of hom-objects in $\mathbb{X}$).

This shows that for example Cat has a factorization which is preserved by products which is inherited from the standard surjective/injective factorization of Set. The factorization is in fact into functors which are surjective on the homsets – the objects do not change but the homsets are quotiented – followed by a faithful functor.

This in turn means that Cat-enriched categories have a factorization ...

3.2.3 Regular and exact categories

A map is **regular epic** if it is the coequalizer of a parallel pair of maps. As before for any map f there is a universal such pair of maps, called the **kernel pair** of f formed by pulling back f along itself. Thus, one can test whether a map is regular by determining whether it is the coequalizer of its kernel pair.

A **regular category** is a category which has pullbacks, coequalizers of all kernel pairs, and has pullbacks of a regular epic along any map regular epic. regular categories where introduced by Michael Barr to capture certain set-like properties. The category of Set is therefore the prototypical regular category. The fact that Set is regular implies that, inparticular, categories of algebras are regular providing a large source of examples of regular categories in the body of traditional mathematics.

It is a classic result of Michael Barr that every regular category admit a factorization system where $\mathbb{E}$ is the class of regular epics and $\mathcal{M}$ is the class of monics. In particular this means that in a regular category the extremal epics are precisely the regular epics.

Proposition 3.2.7 *In any regular category the regular epics and the monics provide a stable factorization system.*

Proof: As monics are closed to composition and contain the isomorphisms we may use proposition 1.4.5 *(iii)* to show that these give a factorization system provided we can show that there is a maximal cofactorization for every map into a regular epic followed by a monic. Now we have already seen (in the dual) that regular epics are orthogonal to monics and this means that the orthogonallity condition of being a maximal factorization will be automatic. Thus, all that really needs to be done is to show that we can factor any map into a regular epic followed by a monic.

Given any map $h : A \longrightarrow B$ we may form its kernel pair and take the coequalizer of that pair. As the original map coequalizes the kernel pair we certainly obtain a factorization:

$$h \wedge h \underset{\pi_1}{\overset{\pi_0}{\rightrightarrows}} A \xrightarrow{e_h} A' \xrightarrow{j} B$$

with $h = e_h j$ where e_h is regular epic. The difficulty is to show that j is monic. This is the special property of regular categories.

Notice that j is monic if and only if its kernel pair is equivalent to $(1_{A'}, 1_{A'})$ so that it would seem sensible to examine the kernel pair of j. The following lemma is then a crucial observation.

Lemma 3.2.8 *In a regular category, if f is regular epic, then the induced comparison map from the kernel pair of fg, for any map g, to that of g is epic.*

PROOF: We have the following diagram of pullbacks:

$$\begin{tikzcd}
(fg)\wedge(fg) \arrow[r, two heads, "f'_1"] \arrow[d, "f'_0"'] \arrow[dr, dotted, "k"] & g\wedge(fg) \arrow[r, "g'_1"] \arrow[d, "f_0"] & A \arrow[d, "f"] \\
(fg)\wedge g \arrow[r, two heads, "f_1"] \arrow[d, "g'_0"'] & g\wedge g \arrow[r, "g_1"] \arrow[d, "g_0"] & B \arrow[d, "g"] \\
A \arrow[r, two heads, "f"'] & B \arrow[r, "g"'] & C
\end{tikzcd}$$

in which the regular epics are indicated by double headed arrows. The comparison maps between the kernel pairs is given by $f = f_1'f_0 = f_0'f_1$ and is epic. □

Setting $f = e_h$ and $j = g$ in the above lemma we obtain $kg_0 = f_0'g_0'f = f_1'g_1'f = kg_1$ so that $g_0 = g_1$ showing that the kernel of j is trivial. □

A, not so obvious, corollary of this is:

Corollary 3.2.9 *In a regular category the composition of two regular epics is a regular epic.*

A factorization is stable in the pullbacks along any map of an $\mathcal{E}$-map is an $\mathcal{E}$-map. We have the following surprisingly strong result:

Proposition 3.2.10 *A stable factorization on a category $\mathbb{X}$ which has all finite limits is the regular epic/monic factorization if and only if $\mathcal{M}$ is precisely the the class of monics.*

PROOF: If it is a regular epic/monic factorization then certainly the $\mathcal{M}$-maps are precisely the monics. Conversely we note that if every monic is in $\mathcal{M}$ then the diagonal maps must be in $\mathcal{M}$ and so each $\mathcal{E}$-map is epic and every monic $\mathcal{E}$-map is an isomorphism. We must show that every $\mathcal{E}$-map is regular epic.

Let $f : A \longrightarrow B$ be an $\mathcal{E}$-map and consider its kernel, E_f given by the pullback

$$\begin{tikzcd}
 & A \arrow[dr, "f"] & \\
E_f \arrow[ur, "\pi_0"] \arrow[dr, "\pi_1"'] & & B \\
 & A \arrow[ur, "f"'] &
\end{tikzcd}$$

We would like

$$E_f \overset{\pi_0^f}{\underset{\pi_1^f}{\rightrightarrows}} A \xrightarrow{f} B$$

to be a coequalizer. To show this is so consider a map g which coequalizes π_0^f, π_1^f

$$E_f \overset{\pi_0^f}{\underset{\pi_1^f}{\rightrightarrows}} A \xrightarrow{g} C$$

then we have

$$E_f \overset{\pi_0^f}{\underset{\pi_1^f}{\rightrightarrows}} A \xrightarrow{\langle f,g\rangle} B \times C$$

and we can factorize $\langle f,g\rangle = e(\langle f,g\rangle)m(\langle f,g\rangle)$ and

$$E_f \overset{\pi_0^f}{\underset{\pi_1^f}{\rightrightarrows}} A \xrightarrow{e(\langle f,g\rangle)} D \xrightarrow{m(\langle f,g\rangle)} B \times C$$

where as $m(\langle f,g\rangle)$ is monic $e(\langle f,g\rangle)$ coequalizes the maps but now we have comparison maps $m(\langle f,g\rangle)\pi_0 : D \longrightarrow B$ and $m(\langle f,g\rangle)\pi_1 : D \longrightarrow C$ where the former is in $\mathcal{E}$ as f is. To show that f is the coequalizer it suffices to prove that the first of these maps is monic as then it will be a monic $\mathcal{E}$-map and so an isomorphism providing the required comparison map $(m(\langle f,g\rangle)\pi_0)^{-1}m(\langle f,g\rangle)\pi_1 : B \longrightarrow C$.

Suppose therefore $xm(\langle f,g\rangle)\pi_0 = ym(\langle f,g\rangle)\pi_0$ then we have, by forming the pullback squares below

$$\begin{tikzcd}
 & & A \arrow[dr, "e(\langle f,g\rangle)"] \arrow[ddrr, bend left, "f"] & & \\
 & X' \arrow[ur, two heads, "x'"] \arrow[dr, "\alpha'"] & & D \arrow[dr, "m(\langle f,g\rangle)\pi_0"] & \\
Z \arrow[ur, two heads, "\beta"] \arrow[dr, two heads, "\alpha"'] & & X \arrow[ur, "x"] \arrow[dr, "y"'] & & B \\
 & Y' \arrow[ur, two heads, "\beta'"] \arrow[dr, "y'"'] & & D \arrow[ur, "m(\langle f,g\rangle)\pi_0"'] & \\
 & & A \arrow[ur, two heads, "e(\langle f,g\rangle)"] \arrow[uurr, bend right, "f"'] & &
\end{tikzcd}$$

an $\mathcal{E}$-map (and therefore an epic map) $k = \alpha\beta' : Z \longrightarrow X$ and also, as the outer square commutes a map $h : X \longrightarrow E_f$ such that $\beta x' = h\pi_0^f$ and $\alpha y' = h\pi_1^f$ so that

$$kx = \beta\alpha' x = \beta x' e(\langle f,g\rangle) = h\pi_0^f e(\langle f,g\rangle) = h\pi_1^f e(\langle f,g\rangle) = \alpha y' e(\langle f,g\rangle) = ky$$

so as k is epic $x = y$ and $m(\langle f,g\rangle)\pi_0$ is monic. $\square$

We may define an equivalence relation on A, (E, π_0, π_1) in a category with pullbacks to be a jointly monic pair:

$$\begin{tikzcd}
 & E \arrow[dl, "\pi_0"'] \arrow[dr, "\pi_1"] & \\
A & & A
\end{tikzcd}$$

which is

- Reflexive: that is there is a map $\Delta : A \longrightarrow E$ with $\Delta\pi_0 = 1_A = \Delta\pi_1$;

- Symmetric: that is there is a map $c : E \longrightarrow E$ such that $c\pi_0 = \pi_1$ and $c\pi_1 = \pi_0$ (notice that this implies that $cc = 1_E$ as π_0 and π_1 are jointly monic);
- Transitive: there is a map $t : E_2 \longrightarrow E$ where

$$
\begin{tikzcd}
E_2 \arrow[r, "\pi_1'"] \arrow[d, "\pi_0'"'] & E \arrow[d, "\pi_0"] \\
E \arrow[r, "\pi_1"'] & A
\end{tikzcd}
$$

  is a pullback and where $t\pi_0 = \pi_0'\pi_0$ and $t\pi_1 = \pi_1'\pi_1$ (notice that t is uniquely determined as the pair π_0, π_1 is jointly monic).

It is not hard to see that any kernel pair forms an equivalence relation in this purely formal sense:

Lemma 3.2.11 *The kernel pair of any map is an equivalence relation.*

PROOF: We define the required maps by:

$$
\begin{tikzcd}
& & A \arrow[dr, "f"] & \\
A \arrow[urr, equal] \arrow[drr, equal] \arrow[r, dotted, "\Delta"] & E \arrow[ur, "\pi_0"] \arrow[dr, "\pi_1"'] & & B \\
& & A \arrow[ur, "f"'] &
\end{tikzcd}
\qquad
\begin{tikzcd}
& & A \arrow[dr, "f"] & \\
E \arrow[urr, "\pi_1"] \arrow[drr, "\pi_0"'] \arrow[r, dotted, "c"] & E \arrow[ur, "\pi_0"] \arrow[dr, "\pi_1"'] & & B \\
& & A \arrow[ur, "f"'] &
\end{tikzcd}
\qquad
\begin{tikzcd}
& & E_2 \arrow[dl] \arrow[d, dotted, "t"] \arrow[dr] & & \\
& E \arrow[dl] \arrow[dr] & E \arrow[dll] \arrow[drr] & E \arrow[dl] \arrow[dr] & \\
A \arrow[drr, "f"'] & & A \arrow[d, "f"] & & A \arrow[dll, "f"] \\
& & A & &
\end{tikzcd}
$$

where $E = f \wedge f$ and E_2 is the pullback of the small upper square at whose apex its stands. $\square$

It is not the case that every equivalence relation in this formal sense will be, in an arbitary regular category, a kernel pair. An **exact category** is precisely a regular category in which every equivalence relation is also a kernel pair. An exact category allows us constructions of structures which are defined by the formation of equivalence relations: these constructions are occur everywhere in mathematics.

The properties of a category relating to the kernel pairs, regular maps, and equivalence relations are referred to as **exactness** properties. These properties are very important in judging how nearly the category behaves like the category Set and indeed there are completeness theorems to the effect that any regular category can be embedded fully and faithfully into a Grothendiek topos (which may be viewed as an abstract set theory) in such a manner as to preserve pullbacks (indeed all finite limits) and coequalizers of kernel pairs.

An exact category need not in general have coequalizers. However, if it is *well-powered* (i.e. every object has only a set of non-isomorpic subobjects) and has limits then one may form coequalizers simply by finding the smallest equivalence relation through which the pair of maps factor and using its coequalizer.

Proposition 3.2.12 *An exact category which has limits and is well-powered has all coequalizers.*

PROOF: The smallest equivalence relation through which a parallel pair (f, g) factors exists as any arbitrary intersection of equivalence relations is, clearly, an equivalence relation and the pair always factors though the chaotic equivalence relation.

Furthermore, the coeqalizer of $f$and g is the coequalizer of this smallest equivalence relation. For, if $fh = gh$ then f and g factor through $\mathsf{Ker}(h)$. This means the smallest such equivalence relation, $E_{(f,g)}$, is contained in $\mathsf{Ker}(h)$ but this gives a unique comparison map from the coequalizer of $E_{(f,g)}$ to $\mathsf{Ker}(h)$. This gives the unique comparison map. $\square$

3.2.4 $\mathcal{M}$-extremal factorizations

In this section we shall provide a slightly surprising link between factorization systems and wide reflexive subcategories. The main result says that (under certain conditions) such reflexive subcategories correspond precisely to reflexive factorization systems. The condition required is that the category is well-powered and has complete subobject lattices: this is a fairly strong condition but it should suggest to the reader that, given a reflexive category, it is sensible to look for a factorization system which is generated by it. Before describing these results we start with some observations on extremal maps.

A **system of monics** $\mathcal{M}$ in a category $\mathbb{X}$ with pullbacks is a class of monics which contains all isomorphism, is closed to composition, and is stable in the sense that the pullback of an $\mathcal{M}$-map along any map is an $\mathcal{M}$-map. If the category $\mathbb{X}$ is well-powered and we say that $\mathcal{M}$ is a **complete system of monics** in case the intersection of an arbitrary set of $\mathcal{M}$-subobjects is an $\mathcal{M}$-subobject.

An **$\mathcal{M}$-extremal map** is a map $f : A \longrightarrow B$ such that any factorization $f = f'm$ through an $\mathcal{M}$-map, m, implies that m is an isomorphism. The monic-extremal maps which are epic are called the **extremal epics**.

Proposition 3.2.13

(i) *In a well-powered category with pullbacks and complete subobject lattices the monics of any class of maps $\mathbb{A} = (\mathbb{A}_\perp)^\perp$ is a complete system of monics.*

(ii) *In a category with pullbacks and a complete system of monics $\mathcal{M}$ the $\mathcal{M}$-extremal maps and the $\mathcal{M}$-subobjects give a factorization system.*

(iii) *In any well-powered category which has equalizers and pullbacks and has complete subobject lattices, the extremal epics and the monics form a factorization system.*

PROOF:

(i) The monics in $\mathbb{A}$ are stable, contain all isomorphism, and are closed to composition as $\mathbb{A}$ has these properties. The only fact we have not mentioned is that the limit of an arbitrary fan of $\mathbb{A}$-maps will be in $\mathbb{A}$. This is left as an exercise.

(ii) First $\mathcal{M}$-extremal epics are precisely the maps which are orthogonal to $\mathcal{M}$. They are orthogonal to $\mathcal{M}$ as consider the square whose top map cannot be factored through a proper

$\mathcal{M}$-subobject and whose bottom map is $\mathcal{M}$.

$$\begin{tikzcd}
A \arrow[rr, "f"] \arrow[dd, "h_1"'] \arrow[dr, dotted, "k"] & & B \arrow[dd, "h_2"] \\
& h_2 \wedge m \arrow[ur, tail, "m'"'] \arrow[dl, "h_2'"] & \\
C \arrow[rr, tail, "m"'] & & D
\end{tikzcd}$$

Form the pullback of m along h_2 inside the square: as f cannot be factored through a proper monic m' must be an isomorphism. This provides the required cross map $m^{-1}h_2' : B \longrightarrow C$ which must be unique as m is monic.

To show that any map which is orthogonal to $\mathcal{M}$ must be $\mathcal{M}$-extremal we consider a map f which can be factorized into $f'm$ where $m \in \mathcal{M}$. This gives the square:

$$\begin{tikzcd}
A \arrow[r, "f"] \arrow[d, "f'"'] & B \arrow[d, equal] \arrow[dl, dotted, "k"'] \\
C \arrow[r, tail, "m"'] & B
\end{tikzcd}$$

from which we immediately obtain that m must be an isomorpism.

This means that both the $\mathbb{E}$-maps and the $\mathcal{M}$-maps contain all isomorphisms and are closed to composition. Furthermore they are orthogonal classes of maps: to show they form a factorization system it, therefore, suffices to show that we may factorize an arbitrary map. Given $f : A \longrightarrow B$, as the $\mathcal{M}$-system is complete, there is a least $\mathcal{M}$-subobject m such that f factors through that subobject as $f'm = f$. Clearly f' must be an $\mathcal{M}$-extremal map as otherwise m would not be the smallest $\mathcal{M}$-subobject.

(iii) Clearly the set of all monics form a complete system of monics when the category has complete subobject lattices. The presence of pullbacks ensures that the monic-extremal maps and the monics form a factorization system. Using corollary 3.2.4, and the fact that the category has equalizers, implies that the monic extremal maps are epic.

□

Of course, these factorizations are not necessarily stable and, certainly, the extremal epics need not be regular epics! However, it is notable that the assumption that the subobject lattices are complete automatically provides a factorization system in the presence of pullbacks. This factorization system explains why extremal epics are of some classical interest.

3.2.5 $\mathcal{M}$-shape subcategories

A wide reflexive subcategory with reflector (R, η) is an $\mathcal{M}$-**shape** subcategory in case each unit η_A is an $\mathbb{E}$-map and every $\mathbb{E}$-map under reflection becomes an isomorphism.

An object is Z is **$\mathbb{E}$-universal** in case given any map $x : X \longrightarrow Z$ and a $\mathbb{E}$-map $g : X \longrightarrow Y$ there is a unique map $x' : YtoZ$ such that $gx' = x$. We may portray the situation as:

$$\begin{tikzcd} X \arrow[r, two heads, "g"] \arrow[dr, "x"'] & Y \arrow[d, dotted, "x'"] \\ & Z \end{tikzcd}$$

A factorization system has **enough $\mathbb{E}$-universal objects** in case every object has an $\mathbb{E}$-map to an $\mathbb{E}$-final object. We have the following observations:

Lemma 3.2.14 *In any category $\mathbb{X}$ with an $(\mathbb{E}, \mathcal{M})$-factorization system:*

(i) *If $\mathbb{X}$ has a final object then $\mathbb{X}$ has enough $\mathbb{E}$-universal objects;*

(ii) *$\mathbb{X}$ has enough $\mathbb{E}$-universal objects if and only if $\mathbb{X}$ has an $\mathcal{M}$-shape subcategory;*

(iii) *The $\mathcal{M}$-shape subcategory if it exists is unique.*

PROOF:

(i) If X has a final object for any object X we may factorize the unique map to the final object:

$$X \xrightarrow{!_X} 1 = X \xrightarrow{\eta_X} R(X) \xrightarrow{!_{R(X)}} 1.$$

I claim the object I have suggestively labelled as $R(X)$ (note if $!_X$ is an $\mathcal{M}$-map we shall let $R(X) = X$ to get a wide reflection) is an $\mathbb{E}$-final object.

To show this we use the factorization system to obtain a unique cross-map for:

$$\begin{tikzcd} Y \arrow[r, two heads, "g"] \arrow[d, "x"'] & Z \arrow[d, "!_Z"] \arrow[dl, dotted, "k"'] \\ R(X) \arrow[r, "!_{R(X)}"'] & 1 \end{tikzcd}$$

Clearly k is unique when both triangles commute but note that the lower triangle *always* commutes so the requirement that the upper triangle commutes secures the uniqueness of k.

(ii) It remains to provide the reflection: set $\eta_X : X \longrightarrow R(X)$ to be a chosen map to an $\mathbb{E}$-final object (being careful to arrange that η_X is the identity in case X is already $\mathbb{E}$-final). Then note that the property of being $\mathbb{E}$-final is precisely the required universal property for a reflection. It remains to check that every $\mathbb{E}$-map g becomes an isomorphism under this reflection: for this we have

$$\begin{tikzcd} A \arrow[r, two heads, "g"] \arrow[d, two heads, "\eta_A"'] & Y \arrow[d, two heads, "\eta_Y"] \\ R(A) & R(Y) \arrow[l, dotted, "(g\eta_Y^\sharp)"] \end{tikzcd}$$

where the inverse is given by the unique map going in the opposite direction.

Conversely if $\mathbb{X}$ has an M-shape subcategory then it suffices to show that each $R(Z)$ is an $\mathbb{E}$-final object. For this we have

$$\begin{tikzcd}
X \arrow[r, two heads, "g"] \arrow[dd, "x"'] \arrow[ddr, "\eta_X"] & Y \arrow[d, "\eta_Y"] \\
 & R(Y) \arrow[d, "R(g)^{-1}"] \\
R(Z) & R(Y) \arrow[l, "R(x)"']
\end{tikzcd}$$

which gives the desired map $\eta_Z R(g)^{-1} R(x)$. For uniqueness suppose k is an alternative map then $k = \eta_Y R(k)$ but $R(k) = R(g)^{-1} R(x)$.

(iii) The $\mathcal{M}$-shape subcategory is always the full subcategory determined by the $\mathbb{E}$-universal objects.

□

3.2.6 Reflexive factorizations and subcategories

Conversely, given a wide reflexive subcategory we wish to construct a factorization system for which the given reflexive subcategory is the $\mathcal{M}$-shape subcategory. We would also like the $\mathbb{E}$-maps to be *precisely* those which become isomorphisms under the reflection. Notice that if this is to be the case then whenever $fg = h$ and any two of f, g and h are $\mathbb{E}$-maps then the third must also be an $\mathbb{E}$-map. This actually means that the factorization must satisfies an additional property, namely, if $fg = h$ and h and g are $\mathbb{E}$-maps then f is an $\mathbb{E}$-map. A **reflexive factorization system** is a factorization system which satisfies this additional property.

Lemma 3.2.15 *In a category with a factorization system an $\mathcal{M}$-shape subcategory exists with reflector R such that $\mathbb{E} = \{g | R(g)$ is iso.$\}$ if and only if the factorization is reflexive with enough $\mathbb{E}$-universal objects.*

PROOF:

($\Rightarrow$) If such a factorization exists it must certainly have enough $\mathbb{E}$-universal objects and be reflexive.

($\Leftarrow$) If the factorization is reflexive and it has enough $\mathbb{E}$-universal objects then a reflector exist. Furthermore, if $R(g)$ is an isomorphism then $g\eta_B = \eta_A R(g)$ so that $g \in \mathbb{E}$.

□

Finally we wish to address the question of when reflexive subcategories correspond to reflexive factorizations:

Theorem 3.2.16 *In a well-powered category with pullbacks in which all subobject lattices are complete every reflexive subcategory is the $\mathcal{M}$-shape subcategory of a reflexive factorization.*

PROOF: We need to provide a reflexive factorization when one is given the reflexive subcategory. In this situation the $\mathbb{E}$-maps are required to be $\mathbb{E} = \{g | R(g) \text{ is iso.}\}$ so that the $\mathcal{M}$ maps must be $\mathbb{E}_\perp$. It suffices therefore to provide a factorization of an arbitrary map.

Suppose therefore that $f : A \longrightarrow B$ then we have by inscribing a pullback inside the naturallity square:

$$\begin{tikzcd}
A \arrow[dr, dotted, "\gamma"] \arrow[drr, bend left, "f"] \arrow[ddr, bend right, "\eta_A"'] & & \\
& \eta_B \wedge R(f) \arrow[r, "\pi_0"'] \arrow[d, "\pi_1"] & B \arrow[d, "\eta_B"] \\
& R(A) \arrow[r, "R(f)"'] & R(B)
\end{tikzcd}$$

where we know that $R(f) \in \mathcal{M}$ and so $\pi_0 \in \mathcal{M}$. However, it is quite possible that $\gamma \notin \mathcal{M}$. To overcome this, notice by proposition 3.2.13 *(i)* that the monic $\mathcal{M}$-maps form a complete system of monics so that they give rise to a factorization system. We factor γ into an $\mathcal{M}$-monic-extremal followed by an $\mathcal{M}$-monic $\gamma = \gamma' m$. Our aim then is to show that $R(\gamma')$ is an isomorphism.

First note that $R(\gamma')$ is certainly a section as $\gamma' m \pi_1 = \eta_B$ so $R(\gamma')R(m\pi_1) = R(\eta_B) = 1_{R(B)}$. But this means by inscribing a pullback inside the naturallity square

$$\begin{tikzcd}
A \arrow[dr, dotted] \arrow[drr, bend left, "\gamma'"] \arrow[ddr, bend right, "\eta_A"'] & & \\
& \eta_Q \wedge R(\gamma') \arrow[r, "p"'] \arrow[d, "q"] & Q \arrow[d, "\eta_Q"'] \\
& R(A) \arrow[r, "R(\gamma')"'] & R(Q)
\end{tikzcd}$$

that p is a monic $\mathcal{M}$-map and so, as γ' is monic $\mathcal{M}$-extremal, it must be an isomorphism. But this gives $p^{-1}qR(\gamma') = \eta_Q$ and so $R(p^{-1}q)R(\gamma') = R(\eta_Q) = 1_{R(Q)}$ whence $R(\gamma')$ is a retract and so an isomorphism. $\square$

This means that when the category also has a final object reflexive factorizations and reflexive subcategories correspond bijectively:

Corollary 3.2.17 *In a well-powered finitely complete category with complete subobject lattices reflexive factorizations and reflexive wide subcategories are in bijective correspondence.*

3.2.7 Exercises

1. Prove that a category is regular in case

   - The category has pullbacks;
   - Every arrow has a regular epic monic factorization;
   - The pullback of a regular epic is a regular epic.

2. Give a direct proof that in a regular category the composite of two regular epics is regular epic.

3. Prove that if fg is a regular epimorphism then g is a regular epimorphism.

4. Show that if $\mathbb{A} = (\mathbb{A}_\perp)^\perp$ then the limit of an arbitary fan of $\mathbb{A}$-maps has its map to the apex of the fan an $\mathbb{A}$-map.

5. Say that an object Y in a category $\mathbb{X}$ with an $(\mathbb{E}, \mathcal{M})$-factorization system is **$\mathbb{E}$-extremal** in case every map out of Y is an $\mathcal{M}$-map. Show that if $\mathbb{X}$ has pushouts then any $\mathbb{E}$-extremal object is an $\mathbb{E}$-final object. Conclude that such a category necessarily has an $\mathcal{M}$-shape subcategory which is determined by the $\mathbb{E}$-extremal objects.

6. (Harder) Show that any finitely complete category with a stable extremal epic monic factorization is a regular category.

3.3 Adjoints, monads, and limits

In this section we shall explore in more detail the interactions between limits and adjoints. This takes us in two directions. First to show the circumstances in which preservation of limits is equivalent to being a right adjoint: this is the content of Freyd's adjoint functor theorem. Second, to investigate the limit and colimit properties of the Eilenberg-Moore category of algebras. This discussion leads us into Beck's tripleability theorems.

3.3.1 Adjoint functor theorems

We have discovered that a right adjoint preserves limits. Peter Freyd proved that under certain conditions the converse is also true: namely if a functor preserves limits then it is a right adjoint. This sort of result result is known as an "adjoint functor theorem" and there are many variants. The importance of these theorems is that whenever you encounter a functor which preserves limits (colimits) you should immediately be suspicious that it does so because it is a right (left) adjoint!

We start by observing that:

Lemma 3.3.1

(i) $G : \mathbb{Y} \longrightarrow \mathbb{X}$ *has a universal pair at* A *if and only if the comma category* A/G *has an initial object;*

(ii) *If* G *preserves limits and* $\mathbb{Y}$ *has a* $\mathcal{G}$*-shaped limit then* A/G *has* $\mathcal{G}$*-shaped limits.*

PROOF:

(i) This is a restatement of the universal property.

(ii) The limit of a $\mathcal{G}$-shaped diagram D in A/G is a cone over the diagram $D; \Pi_1; G,$, where $\Pi_1 : A/G \longrightarrow \mathbb{Y}$ is the projection. As G preserves $\mathcal{G}$-shaped limits this is given by taking the limit $\varprojlim(D; \Pi_1)$ in Y and applying G.

□

This means we may concentrate our efforts on finding an initial object in the presence of limits. By a **weak initial object** we shall mean an object which has a (not necessarily) unique map to each object. By a **weak initial family of objects** we shall mean a set of objects such that every object has a map from one of the objects in the set. We observe:

Lemma 3.3.2 (i) *If W is a weak initial object then any W' with a map $h : W' \longrightarrow W$ is a weak initial object.*

(ii) *If the category $\mathbb{X}$ has products then the category has a weak initial family of objects if any only if $\mathbb{X}$ has a weak initial object.*

(iii) *If $\mathbb{X}$ has equalizers (or pullbacks) and a weak initial object W whose only endo-map is 1_W then W is initial.*

(iv) *If $\mathbb{X}$ has a weak initial object W, has equalizers (or pullbacks), and, in particular, has a monic equalizer $w' : W' \longrightarrow W$ for the set of all W-endomorphisms then W' is initial.*

PROOF:

(i) Immediate.

(ii) The product of the weak initial family is clearly a weak initial object.

(iii) If $\mathbb{X}$ has equalizers and a weak initial object W and f and g are two maps to some object X from W then we may form the equalizer:

$$E \xrightarrow{e} W \underset{g}{\overset{f}{\rightrightarrows}} X.$$

As W is a weak initial object there is a map $h : W \longrightarrow E$ and now $he : W \longrightarrow W$. By assumption this is the identity map so $f = g$. A similar argument work for pullbacks.

(iv) We must show that W' has only the identity map as an endo-map. First we show $w' : W' \longrightarrow W$ is a section. As W is weak initial there is a map $w : W \longrightarrow W'$, now observe $w'ww' = w'$ as w' equalizes all endo-maps of W (and so, in particular, ww' and 1_W) thus, as w' is monic $w'w = 1_W$.

Now suppose that W' has an endo-map $v : W' \longrightarrow W'$ then we obtain an endo-map $wvw' : W \longrightarrow W$: by assumption w' equalizes this and the identity on W so that $w'wvw' = w'$ but $w'wv = v$ so $vw' = w'$ and now as w' is monic we must have $v = 1_{W'}$.

So W' is initial.

□

This gives us almost immediately Freyd's adjoint functor theorem:

Theorem 3.3.3 (Freyd's adjoint functor theorem) *If $\mathbb{X}$ and $\mathbb{Y}$ are* Set*-enriched categories and $\mathbb{Y}$ is complete and $G : \mathbb{Y} \longrightarrow \mathbb{X}$ is a functor which preserves all limits then G has a left adjoint if and only if for each $A \in \mathbb{X}$ the slice category A/G has a weak initial family.*

PROOF: Under these assumptions each A/G is a complete category with a weak initial family. It therefore has a weak initial object. Taking the equalizer of all endo-maps of this object yields an initial object for A/G, which in turn yields the left-adjoint. Conversely, if there is an adjoint then each of these categories has an initial object which can also serves as a one element weak initial family. $\square$

The condition that A/G has a weak initial family is often called the **solution set condition**: more concretely it says that there is a family (set) of arrows $\{r_i : A \longrightarrow G(R_i) | i \in I\}$ such that for any arrow $f : A \longrightarrow G(X)$ there is an $i \in I$ with an arrow $h : R_i \longrightarrow X$ such that

$$\begin{tikzcd} A \arrow[r, "r_i"] \arrow[dr, "f"'] & G(R_i) \arrow[d, "G(h)"] \\ & G(X) \end{tikzcd}$$

so that it is a weak family of universal pairs!

Notice that if $\mathbb{Y}$ is a small category (the objects are a set) this mean that any G which preserves all limits must have an adjoint. Unfortunately, a small complete category is a preorder (see exercises)! So this is not such an interesting use of the theorem as might at first be thought. However, this does completely characterize adjoints from complete posets!

A more significant example of an application (which unfortunately is also more mathematical) is as follows: any category of finitary algebras over Set is complete (groups, rings, Ω-algebras (see later)) over sets then the underlying functor U always preserves limits. A map $X \longrightarrow U(A)$ picks out a set of elements of the algebra A, as the algebras are finitary the cardinallity of the algebra generated by these elements is itself bounded by a cardinal. Thus, these are a small family of maps $X \longrightarrow U(A)$ for which A is generated by the elements picked out by X. These maps form a solutions set. Thus there are always adjoints for the underlying functors for categories of finitary algebras.

There is another version of Freyd's adjoint functor theorem, often called the "special adjoint functor theorem" which we now develop.

A set of objects Γ in a category is said to be a **cogenerating family** if for each parallel pair of distinct arrows

$$A \underset{g}{\overset{f}{\rightrightarrows}} B$$

there is a map $h : B \longrightarrow G$ where $G \in \Gamma$ such that $fh \neq gh$. We say that f and g are **codistinguished** by Γ. If Γ consists of exactly one object G then G is called a **cogenerator**.

Dual to this notion is that of a **generating family**: in this case any pair of unequal parallel arrows can be distinguished by a map from an object in the generating family.

Notice that in any small category the set of all the objects will always form both a generating and cogenerating family. In Set the one element set is a generating object while the two element set is a cogenerating object.

More generally in $\mathsf{Set}^{\mathbb{C}^{\mathrm{op}}}$ if we have two parallel natural transformations

$$F \underset{\beta}{\overset{\alpha}{\rightrightarrows}} G : \mathbb{C}^{\mathrm{op}} \longrightarrow \mathsf{Set}$$

which are unequal then there must be an object $C \in \mathbb{C}$ at which they differ so that

$$F(C) \underset{\beta_C}{\overset{\alpha_C}{\rightrightarrows}} G(C)$$

are distinct set maps. But his means there is an element $c \in F(C)$ at which $\alpha_C(c) \neq \beta_C$. This in turn means that the natural transformation given by the Yoneda lemma

$$\mathcal{Y}(c) : \mathbb{C}(\_, C) \Rightarrow F$$

distinguishes the two natural transformations. Finally, this means that the hom-functors of $\mathbb{C}$ form a generating family in $\mathsf{Set}^{\mathbb{C}^{\mathrm{op}}}$.

Thus, we do have a number of examples of large categories with generating families.

Lemma 3.3.4 *In $\mathbb{X}$ is a Set-enriched category with (set indexed) products then Γ is a cogenerating family if and only if each object A is a subobject of a product of objects in Γ.*

PROOF: We may form the product of cogenerators indexed by the maps to the cogenerators from A. This is monic as any parallel arrows to A can be codistinguished by one of the maps to a cogenerator.

Conversely if $a : A \longrightarrow \prod_{i \in I} G_i$ is a subobject of a product of objects in Γ then then any pair of parallel maps into A is distinguished by the map a. But this means for some $i \in I$ $a\pi_i : A \longrightarrow G_i$ codistinguishes the maps. $\Box$

A category is said to be **well-powered** if for each object the subobject category is equivalent to a small poset (so the objects form a set). Notice that if the category is complete this will mean that each object has a smallest subobject (namely the meet of all the subobjects).

Theorem 3.3.5 (Special adjoint functor theorem) *Let $\mathbb{X}$ and $\mathbb{Y}$ be Set-enriched categories and $\mathbb{Y}$ be well-powered, complete, and have a cogenerating family, then any limit preserving functor $G : \mathbb{Y} \longrightarrow \mathbb{X}$ has a left adjoint.*

PROOF: If we can show that A/G has a weak initial family we are done. First note that A/G also has a cogenerating family namely the family $\Gamma_A = \{A \xrightarrow{f} G(K) | f \in \mathbb{X}(A, G(K)), K \in \Gamma\}$ Notice that to ensure this is a set we have used the fact that $\mathbb{X}$ is Set-enriched.

Now consider a map $h : A \longrightarrow G(X)$ then X is a subobject of a product of cogenerators $x : X \longrightarrow \prod_{i \in I} K_i$ in $\mathbb{Y}$ thus we have:

$$\begin{tikzcd}
A \arrow[drr, "\langle hG(x)\pi_i \rangle"] \arrow[dr, dotted] \arrow[ddr, "h"'] & & \\
& G(S) \arrow[r, dotted] \arrow[d, dotted] & \prod_{f \in \{hG(x)\pi_i | i \in I\}} \partial_1(f) \arrow[d, "G(<\pi_{hG(x)\pi_i} hG(x)\pi_i>_{i \in I})"] \\
& G(X) \arrow[r, "G(x)"'] & \prod_{i \in I} K_i
\end{tikzcd}$$

where the top right corner is an irredundant product (each component occurs once) of generators of A/G. The pullback S in $\mathbb{Y}$ is then a subobject of the irredundant product through which the map h can be factored. This makes subobjects in $\mathbb{Y}$ which give subobjects in A/G of the irredundant products of cogenerators there, a weak initial family in A/G. $\square$

We have noted $\mathsf{Set}^{\mathbb{C}^{\mathrm{op}}}$ has a set of generators – these are the hom-functors – and is easily checked that it is co-well-powered and cocomplete. Thus, by the dual of the special adjoint functor theorem, any functor which preserves colimits from $\mathsf{Set}^{\mathbb{C}^{\mathrm{op}}}$ (or indeed Set itself) has a right adjoint.

In fact $\mathsf{Set}^{\mathbb{C}^{\mathrm{op}}}$ also has a cogenerating object, is well-powered, and complete. Thus any functor which preserves all limits from this category has a left adjoint.

3.3.2 Limit and colimit properties for monads

As has been mentioned the Kleisli category inherits very few properties from its partent. However, one property that it does inherit is coproducts:

Lemma 3.3.6 *If $\mathbb{T} = (T, \eta, \mu)$ is a monad on a category with coproducts then the Kleisli category $\mathbb{X}_{\mathbb{T}}$ has coproducts.*

PROOF: Given $f : X \longrightarrow T(Z)$ and $gY \longrightarrow T(Z)$ there is a unique map $\langle f|g\rangle : X + Y \longrightarrow T(Z)$. It remains to check only that this map is the copairing in the Kleisli category. However, this is immediate as the coprojections are repectively $b_0\eta : X \longrightarrow T(X+Y)$ and $b_1\eta : Y \longrightarrow T(X+Y)$. $\square$

The Eilenberg-Moore category of algebras inherits many more properties from its parent. In particular, it inherits limits:

Lemma 3.3.7 *If $\mathbb{X}$ has all (finite) limits then for any monad $\mathbb{X}^{\mathbb{T}}$ will have (finite) limits and $U^{\mathbb{T}} : \mathbb{X}^{\mathbb{T}} \longrightarrow \mathbb{X}$ creates limits.*

PROOF: Let $D : \mathbb{D} \longrightarrow \mathbb{X}^{\mathbb{T}}$ be a diagram then define the limit as by the algebra

$$T(\lim_{\leftarrow} D; U) \to \lim_{\leftarrow} D; U$$

given by the cone whose projections are

$$T(\lim_{\leftarrow} D; U) \xrightarrow{T(p_Y)\nu_Y} U(D(Y))$$

This is easily checked to be an algebra.

To say that $U_{\mathbb{T}}$ creates limits is to say that if this functor applied to a diagram D underlies to a diagram which has a limit α then there is a unique limit cone of D which underlies to α. Given the way we have constructed the limits this is immediate. $\square$

We now turn to the question of what colimits are present in the Eilenberg-Moore category. One thing that we know is that as $F^{\mathbb{T}}$ is a left adjoint it preserves all colimits. In particular, therefore, coproducts of free algebras exist (which we may also conclude from the result above concerning the Kleisli category). Our aim is to now to show that every algebra may be seen as a colimit of a diagram of free algebras. As any colimit can be rearranged as a coequalizer between coproducts we shall concentrate on coequalizers.

A parallel pair of arrows $A \overset{d_0}{\underset{d_1}{\rightrightarrows}} B$ is said to be **contractible** in case there is a map t such that $td_0 = 1_B$ and $d_0td_1 = d_1td_1$. A map $B \xrightarrow{d} C$ **contractibly coequalizes** d_0 and d_1

$$A \overset{d_0}{\underset{d_1}{\rightrightarrows}} B \xrightarrow{d} C$$

in case there is a map $s : C \longrightarrow B$ such that $ds = td_1$. We observe:

Lemma 3.3.8 *Given a contractible pair of arrows d_0 and d_1:*

(i) $A \underset{d_0 \;\; d_1}{\rightrightarrows} B \xrightarrow{d} C$ *is a colimit if and only if d contractibly coequalizes d_0 and d_1;*

(ii) *The coegalizer of a contractible pair is an absolute colimit in the sense that it is preserved by all functors.*

PROOF:.

(i) If $d : B \longrightarrow C$ is the coequalizer then as $d_0td_1 = d_1td_1$ there is a unique map $s : C \longrightarrow B$ such that $ds = td_1$. In that case sd has the property that $dsd = td_1d = td_0d = d$ so that $sd = 1_C$ as d is being a coequalizer is epic. Thus d contractibly coequalizes d_0 and d_1.

Conversely, if d contractibly coequalizes the pair and $h : B \longrightarrow X$ is another coequalizing map (that is $d_0h = d - 1h$) then sh is a mediating map as $dsh = td_1h = td_0h = h$. It is unique as d is a retraction.

(ii) Clearly the image of any contractible pair under any functor is contractible. Therefore, its coequalizer (if it exists) must contractibly coequalize, however, clearly the image of a map which contractibly coequalizes the pair will contractibly coequalize the image of the pair!

$\square$

The point of contractible pairs this is:

Corollary 3.3.9 *Every algebra in an Eilenberg-Moore category of algebras is the coequalizer of a contractible pair of maps between free algebras.*

PROOF: Notice that when (X, ν) is an algebra

$$T^2(X) \overset{\mu}{\underset{T(\nu)}{\rightrightarrows}} T(X) \xrightarrow{\nu} X$$

is a contractibly coequalized contractible pair! The unit $\eta_{T(X)}$ is the contraction of the pair while η_X is the section of ν. We note that:

$$\begin{aligned} \eta_{T(X)}\mu &= 1_{T(X)} \\ T(\nu)\eta T(\nu) &= T(\nu)\nu\eta = \mu\nu\eta = \mu\eta T(\nu) \\ \eta_X\nu &= 1_X \\ \nu\eta_X &= \eta_{T(X)}T(\nu) \end{aligned}$$

This means

$$\begin{tikzcd}
T^3(X) \arrow[r, shift left, "T^2(\nu)"] \arrow[r, shift right, "T(\mu)"'] \arrow[d, "\mu"] & \mathbb{T}^2(X) \arrow[r, "T(\nu)"] \arrow[d, "\mu"] & T(X) \arrow[d, "\nu"] \\
T^2(X) \arrow[r, shift left, "T(\nu)"] \arrow[r, shift right, "\mu"'] & T(X) \arrow[r, "\nu"'] & X
\end{tikzcd}$$

serially commutes displaying (X, ν) as the contractible coequalizer of maps between free algebras. □

Observe now that the underlying functor from an Eilenberg-Moore categories always creates all the coequalizers that the functors T and T^2 preserve:

Lemma 3.3.10 *Let $\mathbb{T} = (T, \eta, \mu)$ be a monad on $\mathbb{X}$ then $U^{\mathbb{T}}$ creates coequalizers that T preserves and for which $T^2(h)$, where h is the regular epic, is epic.*

Proof: Explictly what this means is that suppose $X \overset{f}{\underset{g}{\rightrightarrows}} Y \xrightarrow{h} Z$ is a coequalizer which is preserved by T and $T^2(h)$ is epic then, whenever $(X, \nu_X) \overset{f}{\underset{g}{\rightrightarrows}} (Y, \nu_Y)$ are morphisms of $\mathbb{T}$-algebras, there is a unique algebra structure ν_Z on Z such that $h : (Y, \nu_Y) \longrightarrow (Z, \nu_Z)$ is an algebra homomorphism which is the coequalizer of f and g in $\mathbb{X}^{\mathbb{T}}$. Our job is to supply the unique algebra structure ν_Z. To this end consider the diagram:

$$\begin{tikzcd}
T^2(X) \arrow[r, shift left, "T^2(f)"] \arrow[r, shift right, "T^2(g)"'] \arrow[d, shift right, "T(\nu_X)"'] \arrow[d, shift left, "\mu"] & \mathbb{T}^2(Y) \arrow[r, "T^2(h)"] \arrow[d, shift right, "T(\nu_Y)"'] \arrow[d, shift left, "\mu"] & T(Z) \arrow[d, shift right, "T(\nu_Z)"'] \arrow[d, shift left, "\mu"] \\
T(X) \arrow[r, shift left, "T(f)"] \arrow[r, shift right, "T(g)"'] \arrow[d, "\nu_X"] & T(Y) \arrow[r, "T(h)"] \arrow[d, "\nu_Y"] & T(Z) \arrow[d, "\nu_Z"] \\
X \arrow[r, shift left, "f"] \arrow[r, shift right, "g"'] & Y \arrow[r, "h"'] & Z
\end{tikzcd}$$

where ν_Z is uniquely determined as $T(h)$ is the coequalizer of $T(x)$ and $T(y)$. That ν_Z is the equalizer $T(\nu_Z)$ and μ follows because this is a contractible coequalization. Clearly $\eta_Z \nu_Z = 1_Z$ and $\eta_{T(Z)}$ acts as the contraction: finally ν_Z coequalizes μ and $T(\nu)$ as $T^2(h)$ is epic and

$$T^2(h)\mu\nu_Z = \mu T(h)\nu_Z = \mu\nu_Y h = T(\nu_Y)\nu_Y h = T^2(h)T(\nu_Z)\nu_Z.$$

□

In particular, as contractible coequalization and idempotent splitting is absolute we have:

Corollary 3.3.11 *For any monad $\mathbb{T}$, $U^{\mathbb{T}} : \mathbb{X}^{\mathbb{T}} \longrightarrow \mathbb{X}$ creates contractible coequalizers and idempotent splittings.*

Returning for a moment to the category of Set we know that each epic map therein is a retraction and the coequalizer of its kernel pair. It is reasonable to wonder whether these coequalizations are not reflected. In fact, more is true:

Lemma 3.3.12

(i) *In any category and for any idempotent e the pair*

$$A \underset{1_A}{\overset{e}{\rightrightarrows}} A$$

is contractible;

(ii) *In any category with pullbacks whenever f is a retraction*

$$\mathsf{Ker}(f) \underset{k_1}{\overset{k_0}{\rightrightarrows}} A$$

is contractible.

(iii) *A contractible pair $f, g : A \longrightarrow B$, with $tf = 1_B$ and $ftg = gtg$, has a colimit if and only if the idempotent gt splits.*

PROOF:

(i) In the first case 1_A is the contraction for the pair.

(ii) In the second case, letting s be a section of f we have

$$\begin{tikzcd}
A \arrow[dr, dotted, "t"] \arrow[drr, "fs"] \arrow[ddr, equal] & & \\
& \mathsf{Ker}(f) \arrow[r, "k_1"'] \arrow[d, "k_0"] & A \arrow[d, "f"] \\
& A \arrow[r, "f"'] & B
\end{tikzcd}$$

then $tk_0 = 1$ and $k_0tk_1 = k_0fsk_1 = k_1fsk_1 = k_1tK_1$ showing this is contractible.

(iii) For the last part note that tg is an idempotent as $tgtg = tftg = 1tg = tg$. Suppose that h coequalizes f and g so $fh = gh$ then $h = tfh = tgh$ so h coequalizes this idempotent. Conversely, if h coequalizes the idempotent so that $h = tgh$ then $fh = ftgh = gtgh = gh$ so h coequalizes f and g.

□

As kernels are obtained as limits this means that maps which underlie to a retraction are always are regular epic in the algebra category. We have already noted that algebra categories over Set inherit the epic-monic factorization system of sets; now we may observe further observe that these algebra categories must be regular as every map which underlies to an epic has its kernel reflected by $U^{\mathbb{T}}$ and it is a contractible pair and so it is regular. Pulling back along a map preserve epics in Set and by reflection must do so in the algebra category.

Corollary 3.3.13 *All Eilenberg-Moore algebra categories over* Set *are exact.*

Proof: Any equivalence relation in sets is a kernel and so a contractible pair, thus, all equivalence relations in an algebra category underly to contractible pairs and so have coequalizers which are reflected. □

In view of 3.2.12 any Eilengerg-Moore algebra category over Set has coequalizers given by forming the smallest equivalence relation through which the given parallel pair of maps factor and taking its coequalizer. Thus, if we had coproducts the category would be cocomplete. But, supposing the underlying category is cocomplete, means that we do have arbitrary coproducts of free algebras. As every object can obtained as a coequalizer of a parallel pair of arrows betwee free algebras we can move any colimit calculation onto one on free objects. However, this means that we can form arbitray colimits. We therefore have:

Theorem 3.3.14 *(Cocompleteness of algebra categories) In* $\mathbb{X}$ *is a complete and cocomplete well-powered exact category and* $\mathbb{T}$ *is a monad whose functor preserves regular epics (in the sense of preserving their kernel coequalization diagrams as colimit diagrams) then* $\mathbb{X}^{\mathbb{T}}$ *is a complete and cocomplete exact category.*

In particular, all Eilenberg-Moore algebra categories over Set *are all exact, complete, and co-complete.*

Proof: Note that T then presrves kernel coequalization and T^2 preserves regular epic maps and therefore the epicness of coequalizers. Thus, $U^{\mathbb{T}}$ reflects coequalization of kernels and so is an exact category which is well-powered. It therefore has coequalization. As coproducts of free objects exist and colimit calculations can be moved onto these objects we obtain cocompleteness.

Clearly and Set monads satisfies these requirements. □

If $F \dashv G : \mathbb{X} \longrightarrow \mathbb{Y}$ is an adjunction then it has a unique comparison functor $K : \mathbb{Y} \longrightarrow \mathbb{X}^{\mathbb{T}}$ to the Eilenberg-Moore category. If the functor K is an isomorphism then we say that $G : Y \longrightarrow \mathbb{X}$ is **precisely monadic**. This development allows us to observe:

Theorem 3.3.15 *(Beck's precise monadicity theorem)* $F \dashv G : \mathbb{X} \longrightarrow \mathbb{Y}$ *is precisely monadic if and only if* G *creates coequalizers of contractible pairs.*

Proof: If the functor is monadic by the above it will reflect coequalizers of contractible pairs. Conversely, if the functor refects coequalizers of contractible pairs then as each algebra can be expressed as a contractible coequalizer of free objects under K it must come from a unique object.

Thus K is injective on objects. Furthermore, as each such contractible colimit is reflected each algebra must also be present in $\mathbb{Y}$: so K is surjective on objects.

Finally for maps, as each object is a colimit of free objects the maps from an object are determined by the maps from its covering free object. This means we need only consider maps of the form $F(X) \longrightarrow Y$. Adjointness makes these correspond to maps $X \longrightarrow G(Y) = U^{\mathbb{T}}(K(Y))$ which provides the bijection on maps. $\Box$

The precise monadicty theorems can be turned into **crude monadicity** theorems in which one seeks a comparison functor to the algebra category for which the two requirement commute up to a natural isomorphism.

Theorem 3.3.16 *(Beck's crude monadicity theorem)* $F \dashv G : \mathbb{X} \longrightarrow \mathbb{Y}$ *is cudely monadic if and only if* G *reflects coequalizers of contractible pairs.*

PROOF: The requirement of reflection is that if the underlying contractible pair has a colimit then a colimit must exist in $\mathbb{Y}$ (it necessarily is preserved). However, the presence of such coequalizers means every algebra is represented up to isomorphism. $\Box$