0

Introduction to category theory

Introduction to Part 0

In Part 0 we recall the basic background in category theory which may be required in later portions of this book. The reader who is familiar with category theory should certainly skip Part 0, but even the reader who is not is advised to consult it only in addition to standard texts.

Most of the material in Part 0 is standard and may also be found in other books. Therefore, on the whole we shall refrain from making historical remarks. However, our exposition differs from treatments elsewhere in several respects.

Firstly, our exposition is slanted towards readers with some acquaintance with logic. Quite early we introduce the notion of a 'deductive system'. For us, this is just a category without the usual equations between arrows. In particular, we do not insist that a deductive system is freely generated from certain axioms, as is customary in logic. In fact, we really believe that logicians should turn attention to categories, which are deductive systems with suitable equations between proofs.

Secondly, we have summarized some of the main thrusts of category theory in the form of succinct slogans. Most of these are due to Bill Lawvere (whose influence on the development of category theory is difficult to overestimate), even if we do not use his exact words. Slogan V represents the point of view of a series of papers by one of the authors in collaboration with Basil Rattray.

Thirdly, we have emphasized the algebraic or equational nature of many of the systems studied in category theory. Just as groups or rings are algebraic over sets, it has been known for a long time that categories with finite products are equational over graphs. More recently, Albert Burroni made the surprising discovery that categories with equalizers are also algebraic over graphs. We have included this result, without going into his more technical concept of 'graphical algebra'.

In Part 0, as in the rest of this book, we have been rather cavalier about set theoretical foundations. Essentially, we are using Gödel–Bernays, as do

most mathematicians, but occasionally we refer to universes in the sense of Grothendieck. The reason for our lack of enthusiasm in presenting the foundations properly is our belief that mathematics should be based on a version of type theory, a variant of which adequate for arithmetic and analysis is developed in Part II. For a detailed discussion of these foundational questions see Hatcher (1982, Chapter 8.)

1 Categories and functors

In this section we present what our reader is expected to know about category theory. We begin with a rather informal definition.

Definition 1.1. A *concrete category* is a collection of two kinds of entities, called *objects* and *morphisms*. The former are sets which are endowed with some kind of structure, and the latter are mappings, that is, functions from one object to another, in some sense preserving that structure. Among the morphisms, there is attached to each object A the *identity mapping* 1_A: $A \to A$ such that $1_A(a) = a$ for all $a \in A$. Moreover, morphisms $f: A \to B$ and $g: B \to C$ may be *composed* to produce a morphism $gf: A \to C$ such that $(gf)(a) = g(f(a))$ for all $a \in A$. (See also Exercise 2 below.)

Examples of concrete categories abound in mathematics; here are just three:

Example C1. The category of *sets*. Its objects are arbitrary sets and its morphisms are arbitrary mappings. We call this category '**Sets**'.

Example C2. The category of *monoids*. Its objects are monoids, that is, semigroups with unity element, and its morphisms are homomorphisms, that is, mappings which preserve multiplication (the semigroup operation) and the unity element.

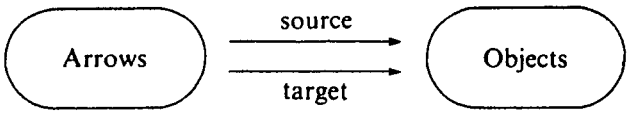
Example C3. The category of *preordered sets*. Its objects are preordered sets, that is, sets with a transitive and reflexive relation on them, and its morphisms are monotone mappings, that is, mappings which preserve this relation.

The reader will be able to think of many other examples: the categories of rings, topological spaces and Banach algebras, to name just a few. In fact, one is tempted to make a generalization, which may be summed up as follows, provided we understand 'object' to mean 'structured set'.

Slogan I. Many objects of interest in mathematics congregate in concrete categories.

We shall now progress from concrete categories to abstract ones, in three easy stages.

Definition 1.2. A *graph* (usually called a *directed graph*) consists of two classes: the class of *arrows* (or *oriented edges*) and the class of *objects* (usually called *nodes* or *vertices*) and two mappings from the class of arrows to the class of objects, called *source* and *target* (often also *domain* and *codomain*).



One writes '$f: A \to B$' for 'source $f = A$ and target $f = B$'. A graph is said to be *small* if the classes of objects and arrows are sets.

Example C4. The category of small *graphs* is another concrete category. Its objects are small graphs and its morphisms are functions F which send arrows to arrows and vertices to vertices so that, whenever $f: A \to B$, then $F(f): F(A) \to F(B)$.

A *deductive system* is a graph in which to each object A there is associated an arrow $1_A: A \to A$, the *identity* arrow, and to each pair of arrows $f: A \to B$ and $g: B \to C$ there is associated an arrow $gf: A \to C$, the *composition* of f with g. A logician may think of the objects as *formulas* and of the arrows as *deductions* or *proofs*, hence of

$$\frac{f: A \to B \quad g: B \to C}{gf: A \to C}$$

as a *rule of inference*. (Deductive systems will be discussed further in Part I.)

A *category* is a deductive system in which the following equations hold, for all $f: A \to B$, $g: B \to C$ and $h: C \to D$:

$$f1_A = f = 1_B f, \quad (hg)f = h(gf).$$

Of course, all concrete categories are categories. A category is said to be *small* if the classes of arrows and objects are sets. While the concrete categories described in examples 1 to 4 are not small, a somewhat surprising observation is summarized as follows:

Slogan II. Many objects of interest to mathematicians are themselves small categories.

Example C1′. Any set can be viewed as a category: a small *discrete*

category. The objects are its elements and there are no arrows except the obligatory identity arrows.

Example C2′. Any monoid can be viewed as a category. There is only one object, which may remain nameless, and the arrows of the monoid are its elements. In particular, the identity arrow is the unity element. Composition is the binary operation of the monoid.

Example C3′. Any preordered set can be viewed as a category. The objects are its elements and, for any pair of objects (a, b), there is at most one arrow $a \to b$, exactly one when $a \leqslant b$.

It follows from slogans I and II that small categories themselves should be the objects of a category worthy of study.

Example C5. The category **Cat** has as objects small categories and as morphisms functors, which we shall now define.

Definition 1.3. A *functor* $F: \mathscr{A} \to \mathscr{B}$ is first of all a morphism of graphs (see Example C4), that is, it sends objects of $\mathscr{A}$ to objects of $\mathscr{B}$ and arrows of $\mathscr{A}$ to arrows of $\mathscr{B}$ such that, if $f: A \to A'$, then $F(f): F(A) \to F(A')$. Moreover, a functor preserves identities and composition; thus

$$F(1_A) = 1_{F(A)}, \quad F(gf) = F(g)F(f).$$

In particular, the identity functor $1_{\mathscr{A}}: \mathscr{A} \to \mathscr{A}$ leaves objects and arrows unchanged and the composition of functors $F: \mathscr{A} \to \mathscr{B}$ and $G: \mathscr{B} \to \mathscr{C}$ is given by

$$(GF)(A) = G(F(A)), \quad (GF)(f) = G(F(f)),$$

for all objects A of $\mathscr{A}$ and all arrows $f: A \to A'$ in $\mathscr{A}$.

The reader will now easily check the following assertion.

Proposition 1.4. When sets, monoids and preordered sets are regarded as small categories, the morphisms between them are the same as the functors between them.

The above definition of a functor $F: \mathscr{A} \to \mathscr{B}$ applies equally well when $\mathscr{A}$ and $\mathscr{B}$ are not necessarily small, provided we allow mappings between classes. Of special interest is the situation when $\mathscr{B} = \mathbf{Sets}$ and $\mathscr{A}$ is small.

Slogan III. Many objects of interest to mathematicians may be viewed as functors from small categories to **Sets**.

Example F1. A set may be viewed as a functor from a discrete one-object category to **Sets**.

Example F2. A small graph may be viewed as a functor from the small category $\cdot \rightrightarrows \cdot$ (with identity arrows not shown) to **Sets**.

Example F3. If $\mathscr{M} = (M, 1, \cdot)$ is a monoid viewed as a one-object category, an $\mathscr{M}$-set may be regarded as a functor from $\mathscr{M}$ to **Sets**. (An $\mathscr{M}$-*set* is a set A together with a mapping $M \times A \to A$, usually denoted by $(m, a) \mapsto ma$, such that $1a = a$ and $(m \cdot m')a = m(m'a)$ for all $a \in A$, m and $m' \in M$.)

Once we admit that functors $\mathscr{A} \to \mathscr{B}$ are interesting objects to study, we should see in them the objects of yet another category. We shall study such functor categories in the next section. For the present, let us mention two other ways of forming new categories from old.

Example C6. From any category (or graph) $\mathscr{A}$ one forms a new category (respectively graph) $\mathscr{A}^{\mathrm{op}}$ with the same objects but with arrows reversed, that is, with the two mappings 'source' and 'target' interchanged. $\mathscr{A}^{\mathrm{op}}$ is called the *opposite* or *dual* of $\mathscr{A}$. A functor from $\mathscr{A}^{\mathrm{op}}$ to $\mathscr{B}$ is often called a *contravariant* functor from $\mathscr{A}$ to $\mathscr{B}$, but we shall avoid this terminology except for occasional emphasis.

Example C7. Given two categories $\mathscr{A}$ and $\mathscr{B}$, one forms a new category $\mathscr{A} \times \mathscr{B}$ whose objects are pairs (A, B), A in $\mathscr{A}$ and B in $\mathscr{B}$, and whose arrows are pairs $(f, g): (A, B) \to (A', B')$, where $f: A \to A'$ in $\mathscr{A}$ and $g: B \to B'$ in $\mathscr{B}$. Composition of arrows is defined componentwise.

Definition 1.5. An arrow $f: A \to B$ in a category is called an *isomorphism* if there is an arrow $g: B \to A$ such that $gf = 1_A$ and $fg = 1_B$. One writes $A \cong B$ to mean that such an isomorphism exists and says that A is *isomorphic* with B.

In particular, a functor $F: \mathscr{A} \to \mathscr{B}$ between two categories is an isomorphism if there is a functor $G: \mathscr{B} \to \mathscr{A}$ such that $GF = 1_{\mathscr{A}}$ and $FG = 1_{\mathscr{B}}$. We also remark that a group is a one-object category in which all arrows are isomorphisms.

To end this section, we shall record three basic isomorphisms. Here **1** is the category with one object and one arrow.

Proposition 1.6. For any categories $\mathscr{A}$, $\mathscr{B}$ and $\mathscr{C}$,

$$\mathscr{A} \times \mathbf{1} \cong \mathscr{A}, \quad (\mathscr{A} \times \mathscr{B}) \times \mathscr{C} \cong \mathscr{A} \times (\mathscr{B} \times \mathscr{C}), \quad \mathscr{A} \times \mathscr{B} \cong \mathscr{B} \times \mathscr{A}.$$

Exercises

1. Prove Propositions 1.4 and 1.6.
2. Show that for any concrete category $\mathscr{A}$ there is a functor $U: \mathscr{A} \to$ **Sets**

which 'forgets' the structure, often called the *forgetful* functor. Clearly U is *faithful* in the sense that, for all $f, g: A \rightrightarrows B$, if $U(f) = U(g)$ then $f = g$. (A more formal version of Definition 1.1 describes a *concrete* category as a pair $(\mathscr{A}, U)$, where $\mathscr{A}$ is a category and $U: \mathscr{A} \to \mathbf{Sets}$ is a faithful functor.)

3. Show that for any category $\mathscr{A}$ there are functors $\Delta: \mathscr{A} \to \mathscr{A} \times \mathscr{A}$ and $\bigcirc_{\mathscr{A}}: \mathscr{A} \to \mathbf{1}$ given on objects A of $\mathscr{A}$ by $\Delta(A) = (A, A)$ and $\bigcirc_{\mathscr{A}}(A) =$ the object of $\mathbf{1}$.

2 Natural transformations

In this section we shall investigate morphisms between functors.

Definition 2.1. Given functors $F, G: \mathscr{A} \rightrightarrows \mathscr{B}$, a *natural transformation* $t: F \to G$ is a family of arrows $t(A): F(A) \to G(A)$ in $\mathscr{B}$, one arrow for each object A of $\mathscr{A}$, such that the following square commutes for all arrows $f: A \to B$ in $\mathscr{A}$:

$$\begin{tikzcd} F(A) \arrow[r, "t(A)"] \arrow[d, "F(f)"'] & G(A) \arrow[d, "G(f)"] \\ F(B) \arrow[r, "t(B)"'] & G(B) \end{tikzcd}$$

that is to say, such that

$$G(f)t(A) = t(B)F(f).$$

It is this concept about which it has been said that it necessitated the invention of category theory. We shall give examples of natural transformations later. For the moment, we are interested in another example of a category.

Example C8. Given categories $\mathscr{A}$ and $\mathscr{B}$, the *functor category* $\mathscr{B}^{\mathscr{A}}$ has as objects functors $F: \mathscr{A} \to \mathscr{B}$ and as arrows natural transformations. The *identity* natural transformation $1_F: F \to F$ is of course given by stipulating that $1_F(A) = 1_{F(A)}$ for each object A of $\mathscr{A}$. If $t: F \to G$ and $u: G \to H$ are natural transformations, their *composition* $u \circ t$ is given by stipulating that $(u \circ t)(A) = u(A)t(A)$ for each object A of $\mathscr{A}$.

To appreciate the usefulness of natural transformations, the reader should prove for himself the following, which supports Slogan III.

Proposition 2.2. When objects such as sets, small graphs and $\mathscr{M}$-sets are

viewed as functors into **Sets** (see Examples F1 to F3 in Section 1), the morphisms between two objects are precisely the natural transformations. Thus, the categories of sets, small graphs and $\mathscr{M}$-sets may be identified with the functor categories $\mathbf{Sets}^{1}$, $\mathbf{Sets}^{\cdot\rightrightarrows}$ and $\mathbf{Sets}^{\mathscr{M}}$ respectively.

Of course, morphisms between sets are mappings, morphisms between graphs were described in Definition 1.3 and morphisms between $\mathscr{M}$-sets are $\mathscr{M}$-homomorphisms. (An *$\mathscr{M}$-homomorphism* $f: A \to B$ between $\mathscr{M}$-sets is a mapping such that $f(ma) = mf(a)$ for all $m \in M$ and $a \in A$.)

We record three more basic isomorphisms in the spirit of Proposition 1.6.

Proposition 2.3. For any categories $\mathscr{A}$, $\mathscr{B}$ and $\mathscr{C}$,

$$\mathscr{A}^{1} \cong \mathscr{A}, \quad \mathscr{C}^{\mathscr{A}\times\mathscr{B}} \cong (\mathscr{C}^{\mathscr{B}})^{\mathscr{A}}, \quad (\mathscr{A}\times\mathscr{B})^{\mathscr{C}} \cong \mathscr{A}^{\mathscr{C}} \times \mathscr{B}^{\mathscr{C}}.$$

We shall leave the lengthy proof of this to the reader. We only mention here the functor $\mathscr{C}^{\mathscr{A}\times\mathscr{B}} \to (\mathscr{C}^{\mathscr{B}})^{\mathscr{A}}$, which will be used later. We describe its action on objects by stipulating that it assigns to a functor $F: \mathscr{A}\times\mathscr{B} \to \mathscr{C}$ the functor $F^*: \mathscr{A} \to \mathscr{C}^{\mathscr{B}}$ which is defined as follows:

For any object A of $\mathscr{A}$, the functor $F^*(A): \mathscr{B} \to \mathscr{C}$ is given by $F^*(A)(B) = F(A, B)$ and $F^*(A)(g) = F(1_A, g)$, for any object B of $\mathscr{B}$ and any arrow $g: B \to B'$ in $\mathscr{B}$.

For any arrow $f: A \to A'$, $F^*(f): F^*(A) \to F^*(A')$ is the natural transformation given by $F^*(f)(B) = F(f, 1_B)$, for all objects B of $\mathscr{B}$.

Finally, to any natural transformation $t: F \to G$ between functors F, G: $\mathscr{A}\times\mathscr{B} \rightrightarrows \mathscr{C}$ we assign the natural transformation $t^*: F^* \to G^*$ which is given by $t^*(A)(B) = t(A, B)$ for all objects A of $\mathscr{A}$ and B of $\mathscr{B}$.

This may be as good a place as any to mention that natural transformations may also be composed with functors.

Definition 2.4. In the situation

$$\mathscr{D} \xrightarrow{L} \mathscr{A} \underset{G}{\overset{F}{\rightrightarrows}} \mathscr{B} \xrightarrow{K} \mathscr{C},$$

if $t: F \to G$ is a natural transformation, one obtains natural transformations $Kt: KF \to KG$ between functors from $\mathscr{A}$ to $\mathscr{C}$ and $tL: FL \to GL$ between functors from $\mathscr{D}$ to $\mathscr{B}$ defined as follows:

$$(Kt)(A) = K(t(A)), \quad (tL)(D) = t(L(D)),$$

for all objects A of $\mathscr{A}$ and D of $\mathscr{D}$.

If $H: \mathscr{A} \to \mathscr{B}$ is another functor and $u: G \to H$ another natural transform-

ation, then the reader will easily check the following 'distributive laws':

$$K(u \circ t) = (Ku) \circ (Kt), \quad (u \circ t)L = (uL) \circ (tL).$$

If we compare Slogans I and III, we are led to ask: which categories may be viewed as categories of functors into **Sets**? In preparation for an answer to that question we need another definition.

Definition 2.5. If A and B are objects of a category $\mathscr{A}$, we denote by $\mathrm{Hom}_{\mathscr{A}}(A, B)$ the class of arrows $A \to B$. (Later, the subscript $\mathscr{A}$ will often be omitted.) If it so happens that $\mathrm{Hom}_{\mathscr{A}}(A, B)$ is a set for all objects A and B, $\mathscr{A}$ is said to be *locally small.*

One purpose of this definition is to describe the following functor.

Example F4. If $\mathscr{A}$ is a locally small category, then there is a functor $\mathrm{Hom}_{\mathscr{A}}: \mathscr{A}^{\mathrm{op}} \times \mathscr{A} \to \mathrm{Sets}$. For an object (A, B) of $\mathscr{A}^{\mathrm{op}} \times \mathscr{A}$, the value of this functor is $\mathrm{Hom}_{\mathscr{A}}(A, B)$, as suggested by the notation. For an arrow $(g, h): (A, B) \to (A', B')$ of $\mathscr{A}^{\mathrm{op}} \times \mathscr{A}$, where $g: A' \to A$ and $h: B \to B'$ in $\mathscr{A}$, $\mathrm{Hom}_{\mathscr{A}}(g, h)$ sends $f \in \mathrm{Hom}_{\mathscr{A}}(A, B)$ to $hfg \in \mathrm{Hom}_{\mathscr{A}}(A', B')$.

Applying the isomorphism $\mathrm{Sets}^{\mathscr{A}^{\mathrm{op}} \times \mathscr{A}} \to (\mathrm{Sets}^{\mathscr{A}})^{\mathscr{A}^{\mathrm{op}}}$ of Proposition 2.3, we obtain a functor $\mathrm{Hom}^*_{\mathscr{A}}: \mathscr{A}^{\mathrm{op}} \to \mathrm{Sets}^{\mathscr{A}}$ and, dually, a functor $\mathrm{Hom}^*_{\mathscr{A}^{\mathrm{op}}}: \mathscr{A} \to \mathrm{Set}^{\mathscr{A}^{\mathrm{op}}}$. We shall see that the latter functor allows us to assert that $\mathscr{A}$ is isomorphic to a 'full' subcategory of $\mathrm{Sets}^{\mathscr{A}^{\mathrm{op}}}$.

Definition 2.6. A *subcategory* $\mathscr{C}$ of a category $\mathscr{B}$ is any category whose class of objects and arrows is contained in the class of objects and arrows of $\mathscr{C}$ respectively and which is closed under the 'operations' source, target, identity and composition. By saying that a subcategory $\mathscr{C}$ of $\mathscr{B}$ is *full* we mean that, for any objects C, C' of $\mathscr{C}$, $\mathrm{Hom}_{\mathscr{C}}(C, C') = \mathrm{Hom}_{\mathscr{B}}(C, C')$.

For example, a proper subgroup of a group is a subcategory which is not full, but the category of Abelian groups is a full subcategory of the category of all groups.

The arrows $F \to G$ in $\mathbf{Sets}^{\mathscr{A}^{\mathrm{op}}}$ are natural transformations. We therefore write $\mathrm{Nat}(F, G)$ in place of $\mathrm{Hom}(F, G)$ in $\mathbf{Sets}^{\mathscr{A}^{\mathrm{op}}}$.

Objects of the latter category are sometimes called 'contravariant' functors from $\mathscr{A}$ to **Sets**. Among them is the functor $h_A \equiv \mathrm{Hom}_{\mathscr{A}}(-, A)$ which sends the object A' of $\mathscr{A}$ onto the set $\mathrm{Hom}_{\mathscr{A}}(A', A)$ and the arrow $f: A' \to A''$ onto the mapping $\mathrm{Hom}_{\mathscr{A}}(f, 1_A): \mathrm{Hom}_{\mathscr{A}}(A'', A) \to \mathrm{Hom}_{\mathscr{A}}(A', A)$.

The following is known as Yoneda's Lemma.

Proposition 2.7. If $\mathscr{A}$ is locally small and $F: \mathscr{A}^{\mathrm{op}} \to \mathbf{Sets}$, then $\mathrm{Nat}(h_A, F)$ is in one-to-one correspondence with $F(A)$.

Proof. If $a \in F(A)$, we obtain a natural transformation $\breve{a}: h_A \to F$ by stipulat-

ing that $\check{a}(B)\colon \mathrm{Hom}_{\mathscr{A}}(B, A) \to F(B)$ sends $g\colon B \to A$ onto $F(g)(a)$. (Note that F is contravariant, so $F(g)\colon F(A) \to F(B)$.)

Conversely, if $t\colon h_A \to F$ is a natural transformation, we obtain the element $t(A)(1_A) \in F(A)$. It is a routine exercise to check that the mappings $a \mapsto \check{a}$ and $t \mapsto t(A)(1_A)$ are inverse to one another.

Definition 2.8. A functor $H\colon \mathscr{A} \to \mathscr{B}$ is said to be *faithful* if the induced mappings $\mathrm{Hom}_{\mathscr{A}}(A, A') \to \mathrm{Hom}_{\mathscr{B}}(H(A), H(A'))$ sending $f\colon A \to A'$ onto $H(f)\colon H(A) \to H(A')$ for all $A', A \in \mathscr{A}$ are injective and *full* if they are surjective. A *full embedding* is a full and faithful functor which is also injective on objects, that is, for which $H(A) = H(A')$ implies $A = A'$.

Corollary 2.9. If $\mathscr{A}$ is locally small, the Yoneda functor $\mathrm{Hom}^*_{\mathscr{A}^{\mathrm{op}}}\colon \mathscr{A} \to \mathbf{Sets}^{\mathscr{A}^{\mathrm{op}}}$ is a full embedding.

Proof. Writing $H \equiv \mathrm{Hom}^*_{\mathscr{A}^{\mathrm{op}}}$, we see that the induced mapping $\mathrm{Hom}(A, A') \to \mathrm{Nat}(H(A), H(A'))$ sends $f\colon A \to A'$ onto the natural transformation $H(f)\colon H(A) \to H(A')$ which, for all objects B of $\mathscr{A}$, gives rise to the mapping $H(f)(B) = \mathrm{Hom}(1_B, f)\colon \mathrm{Hom}(B, A) \to \mathrm{Hom}(B, A')$. Now $f \in H(A')(A)$, hence $\check{f}\colon H(A) \to H(A')$, as defined in the proof of Proposition 2.7, is given by

$$\begin{aligned}\check{f}(B)(g) &= H(A')(g)(f) = \mathrm{Hom}_{\mathscr{A}}(g, 1_{A'})(f)\\ &= fg = \mathrm{Hom}_{\mathscr{A}}(1_B, f)(g) = H(f)(B)(g),\end{aligned}$$

hence $\check{f} = H(f)$. Thus the mapping $f \mapsto H(f)$ is a bijection and so H is full and faithful.

Finally, to show that H is injective on objects, assume $H(A) = H(A')$, then $\mathrm{Hom}(A, A) = \mathrm{Hom}(A, A')$, so A' must be the target of the identity arrow 1_A, thus $A' = A$.

Exercises

1. Prove propositions 2.2 and 2.3.
2. If **2** is the category $\cdot \to \cdot$ (with identity arrows not shown), show that the objects of $\mathscr{A}^2$ are essentially the arrows of $\mathscr{A}$ and that 'source' and 'target' may be viewed as functors $\delta, \delta'\colon \mathscr{A}^2 \rightrightarrows \mathscr{A}$.
3. If $F, G\colon \mathscr{A} \rightrightarrows \mathscr{B}$ are given functors, show that a natural transformation $t\colon F \to G$ is essentially the same as a functor $t\colon \mathscr{A} \to \mathscr{B}^2$ such that $\delta t = F$ and $\delta' t = G$.
4. Show that the isomorphism in Yoneda's Lemma (Proposition 2.7) is natural in both A and F, that is, if $f\colon B \to A$ and $t\colon F \to G$ then the relevant diagrams commute.