II. Categories and Functors

In Chapter I we discussed various algebraic structures (rings, abelian groups, modules) and their appropriate transformations (homomorphisms). We also saw how certain constructions (for example, the formation of $\operatorname{Hom}_A(A,B)$ for given Λ -modules A,B) produced new structures out of given structures. Over and above this we introduced certain "universal" constructions (direct sum, direct product) and suggested that they constituted special cases of a general, and important, procedure. Our objective in this chapter is to establish the appropriate mathematical language for the general description of mathematical systems and of mappings of systems, insofar as that language is applicable to homological algebra.

The language of categories and functors was first introduced by Eilenberg and MacLane [13] to provide a precise description of the processes involved in algebraic topology. Since then an independent mathematical theory has grown up around the basic concepts of the language and today the development, elaboration and application of this theory constitute an extremely active area of mathematical research. It is not our intention to give a treatment of this developing theory; the reader who wishes to pursue the topic of categorical algebra is referred to the texts [6, 18, 35, 37–39] for further reading. Indeed, the reader familiar with the elements of categorical algebra may use this chapter simply as a source of relevant facts, terminology and notation.

1. Categories

To define a category C we must give three pieces of data:

- (1) a class of objects A, B, C, ...,
- (2) to each pair of objects A, B of \mathfrak{C} , a set $\mathfrak{C}(A, B)$ of morphisms from A to B,
 - (3) to each triple of objects A, B, C of C, a law of composition

$$\mathfrak{C}(A, B) \times \mathfrak{C}(B, C) \longrightarrow \mathfrak{C}(A, C)$$
.

Before giving the axioms which a category must satisfy we introduce some auxiliary notation: this should also serve to relate our terminology 1. Categories 41

and notation with ideas which are already very familiar. If $f \in \mathbb{C}(A, B)$ we may think of the morphism f as a generalized "function" from A to B and write

$$f: A \rightarrow B$$
 or $A \xrightarrow{f} B$;

we call f a morphism from the *domain* A to the *codomain* (or *range*) B. The set $\mathfrak{C}(A,B) \times \mathfrak{C}(B,C)$ consists, of course, of pairs (f,g) where $f:A \rightarrow B$, $g:B \rightarrow C$ and we will write the composition of f and g as g = f or, simply, gf. The rationale for this notation (see the Introduction) lies in the fact that if A,B,C are sets and f,g are functions then the composite function from f to f is the function f given by

$$h(a) = g(f(a)), \quad a \in A$$

Thus if the function symbol is written to the *left* of the argument symbol one is naturally led to write h = fg. (Of course it will turn out that sets, functions and function-composition do constitute a category.)

We are now ready to state the *axioms*. The first is really more of a convention, the latter two being much more substantial.

A 1: The sets $\mathfrak{C}(A_1, B_1)$, $\mathfrak{C}(A_2, B_2)$ are disjoint unless $A_1 = A_2$, $B_1 = B_2$.

A 2: Given $f: A \rightarrow B$, $g: B \rightarrow C$, $h: C \rightarrow D$, then

h(gf) = (hg)f (Associative law of composition).

A 3: To each object A there is a morphism $1_A: A \rightarrow A$ such that, for any $f: A \rightarrow B$, $g: C \rightarrow A$,

$$f 1_A = f$$
, $1_A g = g$ (Existence of identities).

It is easy to see that the morphism 1_A is uniquely determined by Axiom A 3. We call 1_A the *identity morphism* of A, and we will often suppress the suffix A, writing simply

$$f1 = f$$
, $1g = g$.

As remarked, and readily verified, the category \mathfrak{S} of sets, functions and function-composition satisfies the axioms. We often refer to the *category* of sets \mathfrak{S} ; indeed, more generally, in describing a category we omit reference to the law of composition when the morphisms are functions and composition is ordinary function-composition (or when, for some other reason, the law of composition is evident), and we even omit reference to the nature of the morphisms if the context, or custom, makes their nature obvious.

A word is necessary about the significance of Axiom A1. Let us consider this axiom in \mathfrak{S} . It is standard practice today to distinguish two functions if their *domains* are distinct, even if they take the same values whenever they are both defined. Thus the sine function $\sin: \mathbb{R} \to \mathbb{R}$

is distinguished from its extension $\sin : \mathbb{C} \to \mathbb{C}$ to the complex field. However, the two functions

$$\sin: \mathbb{R} \to \mathbb{R}$$
, $\sin: \mathbb{R} \to [-1, 1]$

would normally be regarded as the *same* function, although we have assigned to them different *codomains*. However we will see that it is useful – indeed, essential – in homological algebra to distinguish morphisms unless their (explicitly specified) domains *and* codomains coincide.

It is also crucial in topology. Suppose $f_1: X \to Y_1$, $f_2: X \to Y_2$ are two continuous functions which in fact take the same values, i.e., $f_1(x) = f_2(x)$, $x \in X$. Then it may well happen that one of those functions is contractible whereas the other is not. Take, as an example, $X = S^1$, the unit circle in \mathbb{R}^2 , f_1 the embedding of X in \mathbb{R}^2 and f_2 the embedding of X in \mathbb{R}^2 —(0). Then f_1 is contractible, while f_2 is not, so that certainly f_1 and f_2 should be distinguished.

Notice also that the composition gf is only defined if the codomain of f coincides with the domain of g.

We say that a morphism $f: A \rightarrow B$ in $\mathfrak C$ is isomorphic (or invertible) if there exists a morphism $g: B \rightarrow A$ in $\mathfrak C$ such that

$$gf = 1_A$$
, $fg = 1_B$.

It is plain that g is then itself invertible and is uniquely determined by f; we write $g = f^{-1}$, so that

$$(f^{-1})^{-1} = f.$$

It is also plain that the composite of two invertible morphisms is again invertible and thus the relation

$$A \equiv B$$
 if there exists an invertible $f: A \rightarrow B$

(A is isomorphic to B) is an equivalence relation on the objects of the category \mathfrak{C} . This relation has special names in different categories (one-one correspondence of sets, isomorphism of groups, homeomorphism of spaces), but it is important to observe that it is a *categorical* concept.

We now list several examples of categories.

- (a) The category € of sets and functions;
- (b) the category $\mathfrak T$ of topological spaces and continuous functions;
- (c) the category 6 of groups and homomorphisms;
- (d) the category Ab of abelian groups and homomorphisms;
- (e) the category \mathfrak{V}_F of vectorspaces over the field F and linear transformations:
- (f) the category \mathfrak{G}_c of topological groups and continuous homomorphisms;
 - (g) the category \Re of rings and ring-homomorphisms;

1. Categories 43

(h) the category \Re_1 of rings-with-unity-element and ring-homomorphisms preserving unity-element;

(i) the category $\mathfrak{M}_{\Lambda}^{l}$ of left Λ -modules, where Λ is an object of \mathfrak{R}_{1} , and module-homomorphisms;

(j) the category \mathfrak{M}'_{Λ} of right Λ -modules.

Plainly the list could be continued indefinitely. Plainly also each category carries its appropriate notion of invertible morphisms and isomorphic objects. In all the examples given the morphisms are structure-preserving functions; however, it is important to emphasize that the morphisms of a category need not be functions, even when the objects of the category are sets perhaps with additional structure. To give one example, consider the category \mathfrak{T}_h of spaces and homotopy classes of continuous functions. Since the homotopy class of a composite function depends only on the homotopy classes of its factors it is evident that \mathfrak{T}_h is a category – but the morphisms are not themselves functions. Other examples will be found in Exercises 1.1, 1.2.

Returning to our list of examples, we remark that in examples c, d, e, f, g, i, j the category \mathfrak{C} in question possesses an object 0 with the property that, for any object X in \mathfrak{C} , the sets $\mathfrak{C}(X,0)$ and $\mathfrak{C}(0,X)$ both consist of precisely one element.

Thus in $\mathfrak G$ and $\mathfrak A\mathfrak b$ we may take for 0 any one-element group. It is easy to prove that, if $\mathfrak C$ possesses such an object 0, called a *zero object*, then any two such objects are isomorphic and $\mathfrak C(X,Y)$ then possesses a distinguished morphism,

$$X \rightarrow 0 \rightarrow Y$$
,

called the zero morphism and written 0_{XY} . For any $f: W \rightarrow X$, $g: Y \rightarrow Z$ in \mathfrak{C} we have

$$0_{XY}f = 0_{WY}, \quad g0_{XY} = 0_{XZ}.$$

As with the identity morphism, so with the zero morphism 0_{XY} , we will usually suppress the indices and simply write 0. If $\mathfrak C$ possesses zero objects it is called a *category with zero objects*.

If we turn to example (a) of the category \mathfrak{S} then we notice that, given any set X, $\mathfrak{S}(\emptyset, X)$ consists of just one element (where \emptyset is the empty set) and $\mathfrak{S}(X, (p))$ consists of just one element (where (p) is a one-element set). Thus in \mathfrak{S} there is an *initial* object \emptyset and a *terminal* (or *coinitial*) object (p), but no zero object. The reader should have no difficulty in providing precise definitions of initial and terminal objects in a category \mathfrak{C} , and will readily prove that all initial objects in a category \mathfrak{C} are isomorphic and so, too, are all terminal objects.

The final notion we introduce in this section is that of a *subcategory* \mathfrak{C}_0 of a given category \mathfrak{C} . The reader will readily provide the explicit definition; of particular importance among the subcategories of \mathfrak{C} are the *full*

subcategories, that is, those subcategories \mathfrak{C}_0 of \mathfrak{C} such that

$$\mathfrak{C}_0(A,B) = \mathfrak{C}(A,B)$$

for any objects A, B of \mathfrak{C}_0 . For example, \mathfrak{Ab} is a full subcategory of \mathfrak{B} , but \mathfrak{R}_1 is a subcategory of \mathfrak{R} which is not full.

Exercises:

- 1.1. Show how to represent an ordered set as a category. (Hint: Regard the elements a, b, ... of the set as objects in the category, and the instances $a \le b$ of the ordering relation as morphisms $a \rightarrow b$.) Express in categorical language the fact that the ordered set is directed [16]. Show that a subset of an ordered set, with its natural ordering, is a full subcategory.
- **1.2.** Show how to represent a group as a category with a single object, all morphisms being invertible. Show that a subcategory is then precisely a subgroup. When is the subcategory full?
- **1.3.** Show that the category of groups has a generator. (A *generator U* of a category $\mathfrak C$ is an object such that if $f,g:X\to Y$ in $\mathfrak C,f\neq g$, then there exists $u:U\to X$ with $fu\neq gu$.)
- **1.4.** Show that, in the category of groups, there is a one-one correspondence between elements of G and morphisms $\mathbb{Z} \to G$.
- **1.5.** Carry out exercises analogous to Exercises 1.3, 1.4 for the category of sets, the category of spaces, the category of pointed spaces (i.e. each space has a base-point and morphisms are to preserve base-points, see [21]).
- **1.6.** Set out in detail the natural definition of the *Cartesian product* $\mathfrak{C}_1 \times \mathfrak{C}_2$ of two categories $\mathfrak{C}_1, \mathfrak{C}_2$.
- 1.7. Show that if a category has a zero object, then every initial object, and every terminal object, is isomorphic to that zero object. Deduce that the category of sets has no zero object.

2. Functors

Within a category $\mathfrak C$ we have the morphism sets $\mathfrak C(X,Y)$ which serve to establish connections between different objects of the category. Now the language of categories has been developed to delineate the various areas of mathematical theory; thus it is natural that we should wish to be able to describe connections between different categories. We now formulate the notion of a transformation from one category to another. Such a transformation is called a *functor*; thus, precisely, a functor $F:\mathfrak C\to\mathfrak D$ is a rule which associates with every object X of $\mathfrak C$ an object FX of $\mathfrak D$ and with every morphism f in $\mathfrak C(X,Y)$ a morphism Ff in $\mathfrak D(FX,FY)$, subject to the rules

$$F(fg) = (Ff)(Fg), F(1_A) = 1_{FA}.$$
 (2.1)

2. Functors 45

The reader should be reminded, in studying (2.1), of rules governing homomorphisms of familiar algebraic systems. He should also observe that we have evidently the notion of an *identity* functor and of the composition of functors. Composition is associative and we may thus pass to *invertible* functors and *isomorphic* categories.

We now list several examples of functors. The reader will need to establish the necessary facts and complete the descriptions of the functors.

- (a) The embedding of a subcategory \mathfrak{C}_0 in a category \mathfrak{C} is a functor.
- (b) Let G be any group and let G/G' be its abelianized group, i.e. the quotient of G by its commutator subgroup G'. Then $G \mapsto G/G'$ induces the abelianizing functor Abel: $\mathfrak{G} \rightarrow \mathfrak{G}$. Of course this functor may also be regarded as a functor $\mathfrak{G} \rightarrow \mathfrak{Ab}$. This example enables us to exhibit, once more, the importance of being precise about specifying the codomain of a morphism. Consider the groups $G = C_3$, the cyclic group of order 3 generated by t, say, and $H = S_3$, the symmetric group on three symbols. Let $\varphi: G \to H$ be given by $\varphi(t) = (123)$, the cyclic permutation. Let H_0 be the subgroup of H generated by (123) and let $\varphi_0: G \to H_0$ be given by $\varphi_0(t) = (123)$. It may well appear pedantic to distinguish φ_0 from φ but we justify the distinction when we apply the abelianizing functor Abel: $\mathfrak{G} \to \mathfrak{G}$. For plainly Abel(G) = G, Abel $(H_0) = H_0$, Abel $(\varphi_0) = \varphi_0$, which is an isomorphism. On the other hand, H_0 is the commutator subgroup of H, so that $Abel(H) = H/H_0$ and so $Abel(\varphi) = 0$, the constant homomorphism (or zero morphism) $G \rightarrow H/H_0$ ($\cong C_2$). Thus Abel (φ) and Abel (φ_0) are utterly different!
- (c) Let S be a set and let F(S) be the free abelian group on S as basis. This construction yields the *free functor* $F: \mathfrak{S} \to \mathfrak{Alb}$. Similarly there are free functors $\mathfrak{S} \to \mathfrak{G}$, $\mathfrak{S} \to \mathfrak{B}_F$, $\mathfrak{S} \to \mathfrak{M}_A^l$, $\mathfrak{S} \to \mathfrak{M}_A^r$, etc.
- (d) Underlying every topological space there is a set. Thus we get an underlying functor $U:\mathfrak{T}\to\mathfrak{S}$. Similarly there are underlying functors from all the examples (a) to (j) of categories (in Section 1) to \mathfrak{S} . There are also underlying functors $\mathfrak{M}_A^l\to\mathfrak{Alb}, \mathfrak{M}_A^l\to\mathfrak{Alb}, \mathfrak{R}\to\mathfrak{Alb}$, etc., in which some structure is "forgotten" or "thrown away".
- (e) The fundamental group may be regarded as a functor $\pi: \mathfrak{T}^0 \to \mathfrak{G}$, where \mathfrak{T}^0 is the category of spaces-with-base-point (see [21]). It may also be regarded as a functor $\overline{\pi}: \mathfrak{T}^0_h \to \mathfrak{G}$, where the subscript h indicates that the morphisms are to be regarded as (based) homotopy classes of (based) continuous functions. Indeed there is an evident *classifying* functor $Q: \mathfrak{T}^0 \to \mathfrak{T}^0_h$ and then π factors as $\pi = \overline{\pi}Q$.
- (f) Similarly the (singular) homology groups are functors $\mathfrak{T} \to \mathfrak{Ab}$ (or $\mathfrak{T}_h \to \mathfrak{Ab}$).
- (g) We saw in Chapter I how the set $\mathfrak{M}_{A}^{l}(A, B) = \operatorname{Hom}_{A}(A, B)$ may be given the structure of an abelian group. If we hold A fixed and define

 $\mathfrak{M}^{l}_{A}(A, -): \mathfrak{M}^{l}_{A} \rightarrow \mathfrak{Ab}$ by

$$\mathfrak{M}_{A}^{l}(A, -)(B) = \mathfrak{M}_{A}^{l}(A, B),$$

then $\mathfrak{M}_A^l(A, -)$ is a functor. More generally, for any category \mathfrak{C} and object A of \mathfrak{C} , $\mathfrak{C}(A, -)$ is a functor from \mathfrak{C} to \mathfrak{S} . We say that this functor is *represented* by A. It is an important question whether a given functor (usually to \mathfrak{S}) may be represented in this sense by an object of the category.

In viewing the last example the reader will have noted an asymmetry. We have recognized $\mathfrak{M}_A^l(A,-)$ as a functor $\mathfrak{M}_A^l \to \mathfrak{A}\mathfrak{B}$, but if we look at the corresponding construct $\mathfrak{M}_A^l(-,B): \mathfrak{M}_A^l \to \mathfrak{A}\mathfrak{B}$, we see that this is not a functor. For, writing F for $\mathfrak{M}_A^l(-,B)$, then F sends $f:A_1 \to A_2$ to $Ff: FA_2 \to FA_1$. This "reversal of arrows" turns up frequently in applications of categorical ideas and we now formalize the description.

Given any category \mathfrak{C} , we may form a new category $\mathfrak{C}^{\text{opp}}$, the category opposite to \mathfrak{C} . The objects of $\mathfrak{C}^{\text{opp}}$ are precisely those of \mathfrak{C} , but

$$\mathfrak{C}^{\text{opp}}(X, Y) = \mathfrak{C}(Y, X). \tag{2.2}$$

Then the composition in $\mathfrak{C}^{\text{opp}}$ is simply that which follows naturally from (2.2) and the law of composition in \mathfrak{C} . It is trivial to verify that $\mathfrak{C}^{\text{opp}}$ is a category with the same identity morphisms as \mathfrak{C} , and that if \mathfrak{C} has zero objects, then the same objects are zero objects of $\mathfrak{C}^{\text{opp}}$. Moreover,

$$(\mathfrak{C}^{\text{opp}})^{\text{opp}} = \mathfrak{C} . \tag{2.3}$$

Of course the construction of $\mathfrak{C}^{\text{opp}}$ is merely a formal device. However it does enable us to express precisely the *contravariant* nature of $\mathfrak{M}_A^l(-,B)$ or, more generally, $\mathfrak{C}(-,B)$, and to formulate the concept of *categorical duality* (see Section 3).

Thus, given two categories $\mathfrak C$ and $\mathfrak D$ a contravariant functor from $\mathfrak C$ to $\mathfrak D$ is a functor from $\mathfrak C^{\rm opp}$ to $\mathfrak D$. The reader should note that the effective difference between a functor as originally defined (often referred to as a covariant functor) and a contravariant functor is that, for a contravariant functor F from $\mathfrak C$ to $\mathfrak D$, F maps $\mathfrak C(X,Y)$ to $\mathfrak D(FY,FX)$ and (compare (2.1)) F(fg) = F(g) F(f). We give the following examples of contravariant functors.

- (a) $\mathfrak{C}(-,B)$, for B an object in \mathfrak{C} , is a contravariant functor from \mathfrak{C} to \mathfrak{S} . Similarly, $\mathfrak{M}_A^l(-,B)$, $\mathfrak{M}_A^r(-,B)$ are contravariant functors from \mathfrak{M}_A^l , \mathfrak{M}_A^r respectively to \mathfrak{Ab} . We say that these functors are *represented* by B.
- (b) The (singular) cohomology groups are contravariant functors $\mathfrak{T} \to \mathfrak{Ab}$ (or $\mathfrak{T}_h \to \mathfrak{Ab}$).
- (c) Let A be an object of \mathfrak{M}_{A}^{r} and let G be an abelian group. We saw in Section I. 8 how to give $\operatorname{Hom}_{\mathbb{Z}}(A, G)$ the structure of a left Λ -module.

2. Functors 47

 $\operatorname{Hom}_{\mathbb{Z}}(-,G)$ thus appears as a contravariant functor from \mathfrak{M}_{A}^{r} to \mathfrak{M}_{A}^{l} . Further examples will appear as exercises.

Finally we make the following definitions. Recall from Section 1 the notion of a *full* subcategory. Consistent with that definition, we now define a functor $F: \mathfrak{C} \to \mathfrak{D}$ as *full* if F maps $\mathfrak{C}(A, B)$ onto $\mathfrak{D}(FA, FB)$ for all objects A, B in \mathfrak{C} , and as *faithful* if F maps $\mathfrak{C}(A, B)$ injectively to $\mathfrak{D}(FA, FB)$. Finally F is a *full embedding* if F is full and faithful and one-to-one on objects. Notice that then $F(\mathfrak{C})$ is a full subcategory of \mathfrak{D} (in general, $F(\mathfrak{C})$) is not a category at all).

Exercises:

- 2.1. Regarding ordered sets as categories, identify functors from ordered sets to ordered sets, and to an arbitrary category C. Also interpret the opposite category. (See Exercise 1.1.)
- **2.2.** Regarding groups as categories, identify functors from groups to groups. Show that the opposite of a group is isomorphic to the group.
- **2.3.** Show that the center is *not* a functor $\mathfrak{G} \to \mathfrak{G}$ in any obvious way. Let \mathfrak{G}_{epi} be the subcategory of \mathfrak{G} in which the morphisms are the *surjections*. Show that the center is a functor $\mathfrak{G}_{epi} \to \mathfrak{G}$. Is it a functor $\mathfrak{G}_{epi} \to \mathfrak{G}_{epi}$?
- 2.4. Give examples of underlying functors.
- 2.5. Show that the composite of two functors is again a functor. (Discuss both covariant and contravariant functors.)
- **2.6.** Let Φ associate with each commutative unitary ring R the set of its prime ideals. Show that Φ is a contravariant functor from the category of commutative unitary rings to the category of sets. Assign to the set of prime ideals of R the topology in which a base of neighborhoods is given by the sets of prime ideals containing a given ideal J, as J runs through the ideals of R. Show that Φ is then a contravariant functor to \mathfrak{T} .
- 2.7. Let $F: \mathfrak{C}_1 \times \mathfrak{C}_2 \to \mathfrak{D}$ be a functor from the Cartesian product $\mathfrak{C}_1 \times \mathfrak{C}_2$ to the category \mathfrak{D} (see Exercise 1.6). F is then also called a *bifunctor* from $(\mathfrak{C}_1, \mathfrak{C}_2)$ to \mathfrak{D} . Show that, for each $C_1 \in \mathfrak{C}_1$, F determines a functor $F_{C_1}: \mathfrak{C}_2 \to \mathfrak{D}$ and, similarly, for each $C_2 \in \mathfrak{C}_2$, a functor $F_{C_2}: \mathfrak{C}_1 \to \mathfrak{D}$, such that, if $\varphi_1: C_1 \to C_1'$, $\varphi_2: C_2 \to C_2'$, then the diagram

$$F(C_1, C_2) \xrightarrow{F_{C_2}(\varphi_1)} F(C'_1, C_2)$$

$$F_{C_1}(\varphi_2) \downarrow \qquad \qquad \downarrow^{F_{C_1}(\varphi_2)}$$

$$F(C_1, C'_2) \xrightarrow{F_{C_2}(\varphi_1)} F(C'_1, C'_2)$$
(*)

commutes. What is the diagonal of this diagram? Show conversely that if we have functors $F_{C_1}: \mathfrak{C}_2 \to \mathfrak{D}$, $F_{C_2}: \mathfrak{C}_1 \to \mathfrak{D}$, indexed by the objects of $\mathfrak{C}_1, \mathfrak{C}_2$ respectively, such that $F_{C_1}(C_2) = F_{C_2}(C_1)$ and (*) commutes, then these families of functors determine a bifunctor $G: \mathfrak{C}_1 \times \mathfrak{C}_2 \to \mathfrak{D}$ such that $G_{C_1} = F_{C_1}, G_{C_2} = F_{C_2}$.

2.8. Show that $\mathfrak{C}(-,-):\mathfrak{C}^{opp}\times\mathfrak{C}\to\mathfrak{S}$ is a bifunctor.

3. Duality

Our object in this section is to explain informally the duality principle in category theory. We first give an example taken from Section I. 6. We saw there that the *injective* homomorphisms in \mathfrak{M}_A are precisely the *monomorphisms*, i.e. those morphisms μ such that for all α , β

$$\mu\alpha = \mu\beta \Rightarrow \alpha = \beta . \tag{3.1}$$

(The reader familiar with ring theory will notice the formal similarity with right-regularity.) Similarly the *surjective* homomorphisms in \mathfrak{M}_{Λ} are precisely the *epimorphisms* in \mathfrak{M}_{Λ} , i.e. those morphisms ε such that for all α , β

$$\alpha \varepsilon = \beta \varepsilon \Rightarrow \alpha = \beta . \tag{3.2}$$

(The reader will notice that the corresponding concept in ring theory is left-regularity.) Now given any category, we define a monomorphism μ by (3.1) and an epimorphism ε by (3.2). It is then plain that, if ϕ is a morphism in $\mathfrak C$, then ϕ is a monomorphism in $\mathfrak C$ if and only if it is an epimorphism as a morphism of $\mathfrak C^{\text{opp}}$. It then follows from (2.3) that a statement about epimorphisms and monomorphisms which is true in any category must remain true if the prefixes "epi-" and "mono-" are interchanged and "arrows are reversed". Let us take a trivial example. An easy argument establishes the fact that if $\phi \psi$ is monomorphic then ψ is monomorphic. We may thus apply the "duality principle" to infer immediately that if $\psi \phi$ is epimorphic then ψ is epimorphic. Indeed, the two italicized statements are logically equivalent — either stated for $\mathfrak C$ implies the other for $\mathfrak C^{\text{opp}}$. It is superfluous to write down a proof of the second, once the first has been proved.

It is very likely that the reader will come better to appreciate the duality principle after meeting several examples of its applications. Nevertheless we will give a general statement of the principle; this statement will not be sufficiently formal to satisfy the canons of mathematical logic but will, we hope, be intelligible and helpful.

Let us consider a concept \mathscr{C} (like monomorphism) which is meaningful in any category. Since the objects and morphisms of $\mathfrak{C}^{\text{opp}}$ are those of \mathfrak{C} , it makes sense to apply the concept \mathscr{C} to $\mathfrak{C}^{\text{opp}}$ and then to interpret the resulting statement in \mathfrak{C} . This procedure leads to a new concept \mathscr{C}^{opp} which is related to \mathscr{C} by the rule (writing $\mathscr{C}(\mathfrak{C})$ for the concept \mathscr{C} applied to the category \mathfrak{C})

$$\mathscr{C}^{\text{opp}}(\mathfrak{C}) = \mathscr{C}(\mathfrak{C}^{\text{opp}})$$
 for any category \mathfrak{C} .

Thus if $\mathscr C$ is the concept of monomorphism, $\mathscr C^{\mathrm{opp}}$ is the concept of epimorphism (compare (3.1), (3.2)). We may also say that $\mathscr C^{\mathrm{opp}}$ is obtained from $\mathscr C$ by "reversing arrows". This "arrow-reversing" procedure may

3. Duality 49

thus be applied to definitions, axioms, statements, theorems ..., and hence also to proofs. Thus if one shows that a certain theorem \mathcal{F} holds in any category \mathfrak{C} satisfying certain additional axioms A, B, ..., then theorem \mathcal{F}^{opp} holds in any category \mathfrak{C} satisfying axioms $A^{\text{opp}}, B^{\text{opp}}, ...$ In particular if \mathcal{F} holds in any category so does \mathcal{F}^{opp} .

This automatic process of dualizing is clearly extremely useful and convenient and will be much used in the sequel. However, the reader should be clear about the limitations in the scope of the duality principle. Suppose given a statement \mathcal{S}_0 about a particular category \mathfrak{C}_0 , involving concepts $\mathscr{C}_{01}, \ldots, \mathscr{C}_{0k}$ expressed in terms of the objects and morphisms of \mathfrak{C}_0 . For example, \mathfrak{C}_0 may be the category of groups and \mathcal{S}_0 may be the statement "A finite group of odd order is solvable". Now it may be possible to formulate a statement \mathcal{S} about a general category \mathfrak{C} , and concepts $\mathscr{C}_1, \ldots, \mathscr{C}_k$, so that $\mathscr{S}(\mathfrak{C}_0), \mathscr{C}_1(\mathfrak{C}_0), \ldots, \mathscr{C}_k(\mathfrak{C}_0)$ are equivalent to $\mathscr{S}_0, \mathscr{C}_{01}, \ldots, \mathscr{C}_{0k}$ respectively. We may then dualize $\mathscr{S}, \mathscr{C}_1, \ldots, \mathscr{C}_k$, and interpret the resulting statement in the category \mathfrak{C}_0 . Informally we may describe $\mathscr{S}^{\text{opp}}(\mathfrak{C}_0)$ as the dual of \mathscr{S}_0 but two warnings are in order:

(i) The passage from \mathcal{S}_0 to \mathcal{S} is not single-valued; that is, there may well be several statements about a general category which specialize to the given statement \mathcal{S}_0 about the category \mathfrak{C}_0 . Likewise of course, the concepts $\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_k$ may generalize in many different ways.

(ii) Even if \mathscr{S}_0 is provable in \mathfrak{C}_0 , $\mathscr{S}^{\mathrm{opp}}(\mathfrak{C}_0)$ may well be false in \mathfrak{C}_0 . However, if \mathscr{S} is provable, then this constitutes a proof of \mathscr{S}_0 and of $\mathscr{S}^{\mathrm{opp}}(\mathfrak{C}_0)$. (This does not prevent $\mathscr{S}^{\mathrm{opp}}(\mathfrak{C}_0)$ from being vacuous, of course; we cannot guarantee that the dual in this informal sense is always interesting!)

As an example, consider the statement \mathcal{S}_0 "Every Λ -module is the quotient of a projective module". This is a statement about the category $\mathfrak{C}_0 = \mathfrak{M}^l_A$. Now there is a perfectly good concept of a projective object in any category \mathfrak{C} , based on the notion of an epimorphism. Thus (see Section 10) a projective object is an object P with the property that, given φ and ε ,

 $\begin{array}{c}
\theta \\
\downarrow \\
A \xrightarrow{e} B
\end{array}$

with ε epimorphic, there exists θ such that $\varepsilon\theta = \varphi$. We may formulate the statement \mathscr{S} , for any category \mathfrak{C} , which states that, given any object X in \mathfrak{C} there is an epimorphism $\varepsilon: P \to X$ with P projective. Then $\mathscr{S}(\mathfrak{C}_0)$ is our original statement \mathscr{S}_0 . We may now formulate \mathscr{S}^{opp} which asserts that, given any object X in \mathfrak{C} there is a monomorphism $\mu: X \to I$ with I injective (here "injective" is the evident concept dual to "projective"; the reader may easily formulate it explicitly). Then $\mathscr{S}^{\text{opp}}(\mathfrak{C}_0)$ is the statement "Every Λ -module may be embedded in an injective module". Now it

happens (as we proved in Chapter I) that both $\mathscr{S}(\mathfrak{C}_0)$ and $\mathscr{S}^{\text{opp}}(\mathfrak{C}_0)$ are true, but we cannot infer one from the other. For the right to do so would depend on our having a proof of \mathscr{S} – and, in general, \mathscr{S} is false.

We have said that, if \mathscr{S} is provable then, of course, $\mathscr{S}(\mathfrak{C}_0)$ and $\mathscr{S}^{\text{opp}}(\mathfrak{C}_0)$ are deducible. Clearly, though, this is usually too stringent a criterion; in other words, this principle does not permit us to deduce any but the most superficial of propositions about \mathfrak{C}_0 , since it requires some statement to be true in any category. However, as suggested earlier, there is a refinement of the principle that does lead to practical results. Suppose we confine attention to categories satisfying certain conditions Q. Suppose moreover that these conditions are self-dual in the sense that, if any category \mathfrak{C} satisfies Q, so does $\mathfrak{C}^{\text{opp}}$, and suppose further that \mathfrak{C}_0 satisfies conditions Q. Suppose \mathscr{S} is a statement meaningful for any category satisfying Q and suppose that \mathscr{S} may be proved. Then we may infer both $\mathscr{S}(\mathfrak{C}_0)$ and $\mathscr{S}^{\text{opp}}(\mathfrak{C}_0)$. This principle indicates the utility of proving \mathscr{S} for the entire class of categories satisfying Q instead of merely for \mathfrak{C}_0 . We will meet this situation in Section 9 when we come to discuss abelian categories.

Exercises:

- **3.1.** Show that "epimorphic" means "surjective" and that "monomorphic" means "injective"
 - (i) in S, (ii) in T, (iii) in G.
- **3.2.** Show that the inclusion $\mathbb{Z} \subseteq \mathbb{Q}$ is an epimorphism in the category of integral domains. Generalize to other epimorphic non-surjections in this category.
- **3.3.** Consider the underlying functor $U: \mathfrak{T} \to \mathfrak{S}$. Show that $j: X_0 \to X$ in \mathfrak{T} is a homeomorphism of X_0 into X if and only if it is a monomorphism and, for any $f: Y \to X$ in \mathfrak{T} , a factorization $U(j) g_0 = U(f)$ in \mathfrak{S} implies $jf_0 = f$ in \mathfrak{T} with $g_0 = U(f_0)$. Dualize this categorical property of j and obtain a topological characterization of the dual categorical property.
- **3.4.** Define the *kernel* of a morphism $\varphi: A \to B$ in a category with zero morphisms \mathfrak{C} as a morphism $\mu: K \to A$ such that (i) $\varphi \mu = 0$, (ii) if $\varphi \psi = 0$, then $\psi = \mu \psi'$ and ψ' is unique. Identify the kernel, so defined, in \mathfrak{A} b and \mathfrak{G} . Dualize to obtain a definition of *cokernel* in \mathfrak{C} . Identify the cokernel in \mathfrak{A} b and \mathfrak{G} . Let \mathfrak{S}^0 be the category of sets with base points. Identify kernels and cokernels in \mathfrak{S}^0 .
- **3.5.** Generalize the definitions of kernel (and cokernel) above to *equalizers* (and *coequalizers*) of two morphisms $\varphi_1, \varphi_2 : A \rightarrow B$. A morphism $\mu : E \rightarrow A$ is the *equalizer* of φ_1, φ_2 if (i) $\varphi_1 \mu = \varphi_2 \mu$, (ii) if $\varphi_1 \psi = \varphi_2 \psi$ then $\psi = \mu \psi'$ and ψ' is unique. Exhibit the kernel as an equalizer. Dualize.

4. Natural Transformations

We come now to the idea which deserves to be considered the original source of category theory, since it was in the (successful!) attempt to

make precise the notion of a *natural transformation* that Eilenberg and MacLane were led to introduce the language of categories and functors (see [13]).

Let F, G be two functors from the category $\mathfrak D$ to the category $\mathfrak D$. Then a natural transformation t from F to G is a rule assigning to each object X in $\mathfrak C$ a morphism $t_X: FX \to GX$ in $\mathfrak D$ such that, for any morphism $f: X \to Y$ in $\mathfrak C$, the diagram

$$FX \xrightarrow{t_X} GX$$

$$Ff \downarrow \qquad \qquad \downarrow Gf$$

$$FY \xrightarrow{t_Y} GY$$

commutes. If t_X is isomorphic for each X then t is called a natural equivalence and we write $F \simeq G$. It is plain that then $t^{-1}: G \simeq F$, where t^{-1} is given by $(t^{-1})_X = (t_X)^{-1}$. If $t: F \to G$, $u: G \to H$ are natural transformations then we may form the composition $ut: F \to H$, given by $(ut)_X = (u_X)(t_X)$; and the composition of natural transformations is plainly associative. Let $F: \mathfrak{C} \to \mathfrak{D}$, $G: \mathfrak{D} \to \mathfrak{C}$ be functors such that $GF \simeq I: \mathfrak{C} \to \mathfrak{C}$, $FG \simeq I: \mathfrak{D} \to \mathfrak{D}$, where I stands for the identity functor in any category. We then say that \mathfrak{C} and \mathfrak{D} are equivalent categories. Of course, isomorphic categories are equivalent, but equivalent categories need not be isomorphic (see Exercise 4.1). We now give some examples of natural transformations; we draw particular attention to the first example which refers to the first explicitly observed example of a natural transformation.

- (a) Let V be a vector space over the field F, let V^* be the dual vector space and V^{**} the double dual. There is a linear map $\iota_V: V \to V^{**}$ given by $v \mapsto \tilde{v}$ where $\tilde{v}(\varphi) = \varphi(v), v \in V, \varphi \in V^*, \tilde{v} \in V^{**}$. The reader will verify that ι is a natural transformation from the identity functor $I: \mathfrak{B}_F \to \mathfrak{B}_F$ to the double dual functor $**: \mathfrak{B}_F \to \mathfrak{B}_F$. Now let \mathfrak{B}_F^f be the full subcategory of \mathfrak{B}_F consisting of *finite-dimensional* vector spaces. It is then, of course, a basic theorem of linear algebra that ι , restricted to \mathfrak{B}_F^f , is a natural equivalence. (More accurately, the classical theorem says that ι_V is an isomorphism for each V in \mathfrak{B}_F^f .) The proof proceeds by observing that $V \cong V^*$ if V is finite-dimensional. However, this last isomorphism is not natural to define it one needs to choose a basis for V and then to associate with this basis the dual basis of V^* . That is, the isomorphism between V and V^* depends on the choice of basis and lacks the canonical nature of the isomorphism ι_V between V and V^{**} .
- (b) Let G be a group and let G/G' be its commutator factor group. There is an evident surjection $\kappa_G: G \to G/G'$ and κ is a natural transformation from the identity functor $\mathfrak{G} \to \mathfrak{G}$ to the abelianizing functor Abel: $\mathfrak{G} \to \mathfrak{G}$.

- (c) Let A be an abelian group and let A_F be the free abelian group on the set A as basis. There is an evident surjection $\tau_A: A_F \to A$, which maps the basis elements of A_F identically, and τ is a natural transformation from FU to I, where $U: \mathfrak{Alb} \to \mathfrak{S}$ is the underlying functor and $F: \mathfrak{S} \to \mathfrak{Alb}$ is the free functor.
- (d) The Hurewicz homomorphism from homotopy groups to homology groups (see e.g. [21]) may be interpreted as a natural transformation of functors $\mathfrak{T}^0 \to \mathfrak{Ab}$ (or $\mathfrak{T}^0_h \to \mathfrak{Ab}$).

We continue with the following important remark. Given two categories $\mathfrak{C}, \mathfrak{D}$, the reader is certainly tempted to regard the functors $\mathfrak{C} \to \mathfrak{D}$ as the objects of a new category with the natural transformations as morphisms. The one difficulty about this point of view is that it is not clear from a foundational viewpoint that the natural transformations of functors $\mathfrak{C} \to \mathfrak{D}$ form a set. This objection may be circumvented by adopting a set-theoretical foundation different from ours (see [32]) or simply by insisting that the collection of objects of \mathfrak{C} form a set; such a category \mathfrak{C} is called a small category. Thus if \mathfrak{C} is small we may speak of the category of functors (or functor category) from \mathfrak{C} to \mathfrak{D} which we denote by $\mathfrak{D}^{\mathfrak{C}}$ or $[\mathfrak{C}, \mathfrak{D}]$. In keeping with this last notation we will denote the collection of natural transformations from the functor F to the functor G by [F, G].

We illustrate the notion of the category of functors with the following example. Let \mathfrak{C} be the category with two objects and identity morphisms only. A functor $F:\mathfrak{C}\to\mathfrak{D}$ is then simply a pair of objects in \mathfrak{D} , and a natural transformation $t:F\to G$ is a pair of morphisms in \mathfrak{D} . Thus it is seen that $\mathfrak{D}^{\mathfrak{C}}=[\mathfrak{C},\mathfrak{D}]$ is the Cartesian product of the category \mathfrak{D} with itself, that is the category $\mathfrak{D}\times\mathfrak{D}$ in the notation of Exercise 1.6.

We close this section with an important proposition. We have seen that, if A, B are objects of a category \mathfrak{C} , then $\mathfrak{C}(A, -)$ is a (covariant) functor $\mathfrak{C} \to \mathfrak{S}$ and $\mathfrak{C}(-, B)$ is a contravariant functor $\mathfrak{C} \to \mathfrak{S}$. If $\theta : B_1 \to B_2$ let us write θ_* for $\mathfrak{C}(A, \theta_1) : \mathfrak{C}(A, B_1) \to \mathfrak{C}(A, B_2)$, so that

$$\theta_{\star}(\varphi) = \theta \varphi$$
, $\varphi : A \rightarrow B_1$,

and if $\psi: A_2 \longrightarrow A_1$ let us write ψ^* for $\mathfrak{C}(\psi, B): \mathfrak{C}(A_1, B) \longrightarrow \mathfrak{C}(A_2, B)$ so that

$$\psi^*(\varphi) = \varphi \psi$$
, $\varphi: A_1 \longrightarrow B$.

These notational simplifications should help the reader to understand the proof of the following proposition.

Proposition 4.1. Let τ be a natural transformation from the functor $\mathfrak{C}(A,-)$ to the functor F from \mathfrak{C} to \mathfrak{S} . Then $\tau \mapsto \tau_A(1_A)$ sets up a one-one correspondence between the set $[\mathfrak{C}(A,-),F]$ of natural transformations from $\mathfrak{C}(A,-)$ to F and the set F(A).

Proof. We show first that τ is entirely determined by the element $\tau_A(1_A) \in F(A)$. Let $\varphi : A \rightarrow B$ and consider the commutative diagram

$$\begin{array}{ccc}
\mathbb{C}(A,A) \xrightarrow{\varphi_{\star}} \mathbb{C}(A,B) \\
\downarrow^{\tau_{A}} & \downarrow^{\tau_{B}} \\
FA \xrightarrow{F\varphi} FB
\end{array}$$

Then $\tau_B(\varphi) = (\tau_B)(\varphi_*)(1_A) = (F\varphi)(\tau_A)(1_A)$, proving the assertion. The proposition is therefore established if we show that, for any $\kappa \in FA$, the rule

$$\tau_B(\varphi) = (F\varphi)(\kappa), \quad \varphi \in \mathfrak{C}(A, B),$$
(4.1)

does define a natural transformation from $\mathfrak{C}(A,-)$ to F. Let $\theta: B_1 \rightarrow B_2$ and consider the diagram

$$\mathbb{C}(A, B_1) \xrightarrow{\theta_{\star}} \mathbb{C}(A, B_2)$$

$$\downarrow^{\tau_B} \qquad \qquad \downarrow^{\tau_B}$$

$$FB_1 \xrightarrow{F\theta} FB_2$$

We must show that this diagram commutes if τ_{B_1} , τ_{B_2} are defined as in (4.1). Now $(\tau_{B_2}) \theta_*(\varphi) = (\tau_{B_2}) (\theta \varphi) = F(\theta \varphi) (\kappa) = F(\theta) F(\varphi) (\kappa) = F(\theta) \tau_{B_1}(\varphi)$ for $\varphi : A \rightarrow B_1$. Thus the proposition is completely proved. \square

By choosing $F = \mathfrak{C}(A', -)$ we obtain

Corollary 4.2. The set of morphisms $\mathfrak{C}(A',A)$ and the set of natural transformations $[\mathfrak{C}(A,-),\mathfrak{C}(A',-)]$ are in one-to-one correspondence, the correspondence being given by $\psi\mapsto \psi^*, \ \psi:A'\to A$.

Proof. If τ is such a natural transformation, let $\psi = \tau_A(1_A)$, so that $\psi : A' \to A$. Then, by (4.1) τ is given by

$$\tau_B(\varphi) = \varphi_*(\psi) = \varphi \psi = \psi^*(\varphi) \; .$$

Thus $\tau_B = \psi^*$. Of course ψ is uniquely determined by τ and every ψ does induce a natural transformation $\mathfrak{C}(A, -) \rightarrow \mathfrak{C}(A', -)$. Thus the rule $\tau \mapsto \tau_A(1_A)$ sets up a one-one correspondence, which we write $\tau \mapsto \psi$, between the set of natural transformations $\mathfrak{C}(A, -) \rightarrow \mathfrak{C}(A' -)$ and the set $\mathfrak{C}(A', A)$. \square

With respect to the correspondence $\tau \mapsto \psi$ we easily prove

Proposition 4.3. Let $\tau: \mathfrak{C}(A,-) \rightarrow \mathfrak{C}(A',-), \ \tau': \mathfrak{C}(A',-) \rightarrow \mathfrak{C}(A'',-)$. Then if $\tau \mapsto \psi, \ \tau' \mapsto \psi', \ where \ \psi: A' \rightarrow A, \ \psi': A'' \rightarrow A', \ we have$

$$\tau'\tau \mapsto \psi \psi'$$
.

In particular τ is a natural equivalence if and only if ψ is an isomorphism.

Proof.
$$(\tau'\tau)_B = (\tau'_B)(\tau_B) = \psi'^*\psi^* = (\psi\psi')^*$$
.

Proposition 4.1 is often called the *Yoneda lemma*; it has many applications in algebraic topology and, as we shall see, in homological algebra.

If \mathfrak{C} is a small category we may formulate the assertion of Corollary 4.2 in an elegant way in the functor category $\mathfrak{S}^{\mathfrak{C}}$. Then $A \mapsto \mathfrak{C}(A, -)$ is seen to be an embedding (called the *Yoneda embedding*) of $\mathfrak{C}^{\text{opp}}$ in $\mathfrak{S}^{\mathfrak{C}}$; and Corollary 4.2 asserts further that it is a *full* embedding.

Exercises:

- **4.1.** A full subcategory \mathfrak{C}_0 of \mathfrak{C} is said to be a *skeleton* of \mathfrak{C} if, given any object A of \mathfrak{C} , there exists exactly one object A_0 of \mathfrak{C}_0 with $A_0 \cong A$. Show that every skeleton of \mathfrak{C} is equivalent to \mathfrak{C} , and give an example to show that a skeleton of \mathfrak{C} need not be isomorphic to \mathfrak{C} . Are all skeletons of \mathfrak{C} isomorphic?
- **4.2.** Represent the embedding of the commutator subgroup of G in G as a natural transformation.
- **4.3.** Let $F, G: \mathfrak{C} \to \mathfrak{D}, E: \mathfrak{B} \to \mathfrak{C}, H: \mathfrak{D} \to \mathfrak{C}$ be functors, and let $t: F \to G$ be a natural transformation. Show how to define natural transformations $tE: FE \to GE$, and $Ht: HF \to HG$, and show that H(tE) = (Ht)E. Show that tE and Ht are natural equivalences if t is a natural equivalence.
- **4.4.** Let \mathfrak{C} be a category with zero object and kernels. Let $f: A \to B$ in \mathfrak{C} with kernel $k: K \to A$. Then $f_*: \mathfrak{C}(-,A) \to \mathfrak{C}(-,B)$ is a natural transformation of contravariant functors from \mathfrak{C} to \mathfrak{S}_0 , the category of pointed sets. Show that $X \mapsto \ker(f_*)_X$ is a contravariant functor from \mathfrak{C} to \mathfrak{S}_0 which is represented by K, and explain the sense in which k_* is the kernel of f_* .
- 4.5. Carry out an exercise similar to Exercise 4.4 replacing kernels in C by cokernels in C.
- **4.6.** Let \mathfrak{A} be a small category and let $Y: \mathfrak{A} \to [\mathfrak{A}^{\text{opp}}, \mathfrak{S}]$ be the Yoneda embedding $Y(A) = \mathfrak{A}(-, A)$. Let $J: \mathfrak{A} \to \mathfrak{B}$ be a functor. Define $R: \mathfrak{B} \to [\mathfrak{A}^{\text{opp}}, \mathfrak{S}]$ on objects by $R(B) = \mathfrak{B}(J-, B)$. Show how to extend this definition to yield a functor R, and give reasonable conditions under which Y = RJ.
- **4.7.** Let *I* be any set; regard *I* as a category with identity morphisms only. Describe \mathfrak{C}^I . What is \mathfrak{C}^I if *I* is a set with 2 elements?

5. Products and Coproducts; Universal Constructions

The reader was introduced in Section I. 3 to the *universal property* of the direct product of modules. We can now state this property for a general category \mathfrak{C} .

Definition. Let $\{X_i\}$, $i \in I$, be a family of objects of the category $\mathfrak C$ indexed by the set I. Then a product $(X; p_i)$ of the objects X_i is an object X, together with morphisms $p_i: X \to X_i$, called projections, with the universal property: given any object Y and morphisms $f_i: Y \to X_i$, there exists a unique morphism $f = \{f_i\}: Y \to X$ with $p_i f = f_i$.

As we have said, in the category \mathfrak{M}_{Λ} of (left) Λ -modules, we may take for X the direct product of the modules X_i (Section I. 3). In the