Manifolds and Local Structures

A General Theory

Marco Grandis

World Scientific

Published by

World Scientific Publishing Co. Pte. Ltd.
5 Toh Tuck Link, Singapore 596224
USA office: 27 Warren Street, Suite 401-402, Hackensack, NJ 07601
UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

Library of Congress Control Number: 2021003543

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

MANIFOLDS AND LOCAL STRUCTURES
A General Theory

ISBN 978-981-123-399-9 (hardcover)
ISBN 978-981-123-400-2 (ebook for institutions)
ISBN 978-981-123-401-9 (ebook for individuals)

For any available supplementary material, please visit
https://www.worldscientific.com/worldscibooks/10.1142/12199#t=suppl

Desk Editor: Soh Jing Wen

Printed in Singapore

Contents

This page intentionally left blank

Introduction

0.1 Classical manifolds by an external atlas

A smooth manifold is usually defined as a topological space X equipped with a C^∞-*atlas of charts*, indexed by a set I

$$u^i\colon U_i \to X_i \qquad (i \in I). \tag{0.1}$$

A *chart* u^i is a homeomorphism between an open euclidean space U_i (i.e. an open subspace of some space $\mathbb{R}^n$ with euclidean topology) and an open subspace X_i of X. We assume that these open subsets cover X and that every *transition map* (between open euclidean spaces)

$$u^i_j = u_j u^i\colon U_i \dashrightarrow U_j \qquad (i, j \in I), \tag{0.2}$$

is of class C^∞ (i.e. has continuous partial derivatives of any order).

Here $u_j\colon X_j \to U_j$ is the inverse of u^j, and the 'composite' $u_j u^i\colon U_i \dashrightarrow U_j$ (an abuse of notation) is *partially defined* on the open subset $u_i(X_i \cap X_j)$, possibly empty. We shall often distinguish partial mappings by a dot-marked arrow.

The space X is thus locally euclidean, with a locally constant dimension. It is often required to be Hausdorff paracompact, but we drop these conditions, adding them when it is the case.

If Y is also a smooth manifold, with charts $v^h\colon V_h \to Y_h$ $(h \in H)$, a C^∞-*mapping* $f\colon X \to Y$ is a map such that all the *partial* mappings

$$f^i_h = v_h f u^i\colon U_i \dashrightarrow V_h \qquad (i \in I,\ h \in H), \tag{0.3}$$

are of class C^∞. Again, as in (0.2), there is an abuse of notation: we are 'composing' three arrows

$$u^i\colon U_i \to X_i, \qquad f\colon X \to Y, \qquad v_h\colon Y_h \to V_h,$$

which are not consecutive.

This could be fixed with Ehresmann's *pseudoproduct* (recalled below), but we want to work with ordinary composition in categories of partial mappings, along the following line.

(a) We write as $\mathcal{C}$ the category of *topological spaces and partial continuous mappings defined on open subspaces.* A morphism $f\colon X \nrightarrow Y$ is defined on an open subspace $\mathrm{Def}\, f$ of X; for a consecutive morphism $g\colon Y \nrightarrow Z$, the composite $gf\colon X \nrightarrow Z$ is defined on those $x \in X$ (if any) such that $f(x) \in \mathrm{Def}\, g$.

We write as $\mathcal{C}^\infty$ the subcategory of $\mathcal{C}$ of open euclidean spaces and partial C^∞-mappings, defined on open subspaces.

(b) We replace the homeomorphism $u^i\colon U_i \to X_i$ with the topological embedding $u^i\colon U_i \to X$; the latter has a backward morphism $u_i\colon X \nrightarrow U_i$ in $\mathcal{C}$ (defined on X_i), characterised by the relations

$$u^i = u^i u_i u^i, \qquad u_i = u_i u^i u_i, \tag{0.4}$$

which make these morphisms *partial inverse* to each other (as in semigroup theory).

We can now replace the 'illegitimate compositions' of (0.2) and (0.3) with legitimate ones, in $\mathcal{C}$

$$u_j^i = u_j u^i\colon U_i \nrightarrow X \nrightarrow U_j, \qquad v_h f u^i\colon U_i \nrightarrow X \nrightarrow Y \nrightarrow V_h, \tag{0.5}$$

and we can require that these composites (whose domain and codomain are open euclidean spaces) belong to the subcategory $\mathcal{C}^\infty$.

(c) More generally, for each $r \in \mathbb{N} \cup \{\infty, \omega\}$, we write as $\mathcal{C}^r$ the subcategory of $\mathcal{C}$ of *open euclidean spaces and partial* C^r*-mappings defined on open subspaces.* (C^0 means continuous and C^ω means analytic; for $0 < r < \infty$, a C^r-mapping has all continuous partial derivatives of order $\leqslant r$.)

C^r-manifolds are dealt with as above. Topological manifolds correspond to the case $r = 0$ (in which case the transition maps u_j^i automatically belong to $\mathcal{C}^0$).

(d) Categories of partial mappings will generally be denoted by calligraphic letters. The prime example is the category $\mathcal{S}$ of sets and partial mappings.

As a crucial fact, a category of partial mappings has a canonical order: for two partial mappings $f, g\colon X \nrightarrow Y$ the relation $f \leqslant g$, means that f is a restriction of g (with $\mathrm{Def}\, f$ contained in $\mathrm{Def}\, g$). The order is consistent with composition: $\mathcal{S}$, $\mathcal{C}$ and $\mathcal{C}^r$ are *ordered categories.*

The reader may know, or guess, that the 'categories of partial mappings' (like $\mathcal{S}$ or $\mathcal{C}$) we are using can be obtained by a general construction, starting from an 'ordinary' category (like Set or Top) and a suitable subcategory (the embeddings of the definition-sets). This point, dealt with in Section 5.1, plays here a minor role: we prefer to work directly in the relevant categories of partial mappings.

0.2 Intrinsic manifolds on ordered categories

Loosely speaking, it is possible to define a C^r-manifold in an intrinsic way, *inside the category* $\mathcal{C}^r$, as a collection (U_i) of objects, equipped with a family $(u^i_j\colon U_i \to U_j)$ of transition morphisms — a system of instructions specifying how the different charts U_i should be glued together. The gluing will be realised in an external category, namely in $\mathcal{C}$.

More precisely, we define an (intrinsic) *manifold* on the ordered category $\mathcal{C}^r$, indexed by a set I, as a diagram

$$U = ((U_i), (u^i_j))_I$$

in $\mathcal{C}^r$, consisting of objects U_i (the *charts*, for $i \in I$) and morphisms $u^i_j\colon U_i \to U_j$ (the *transition morphisms*, for $i, j \in I$), satisfying three axioms which use the canonical order of the category $\mathcal{C}^r$ (for $i, j, k \in I$):

(i) $u^i_i = 1_{U_i}$ (*identity law*),

(ii) $u^j_k u^i_j \leqslant u^i_k$ (*composition law*, or *triangle inequality*),

(iii) $u^i_j = u^i_j u^j_i u^i_j$ (*symmetry law*).

From a formal point of view, U is a small category enriched on the ordered category $\mathcal{C}^r$, with an additional symmetry condition. The transition mapping u^i_j plays the role of $\mathrm{Hom}(i, j)$, while axiom (ii) replaces the composition mapping $\mathrm{Hom}(i, j) \times \mathrm{Hom}(j, k) \to \mathrm{Hom}(i, k)$.

All this makes sense in the theory of enriched categories on ordered categories, reviewed in Chapter 6.

Plainly, if we start from the usual charts $u^i\colon U_i \to X$ and define their transition morphisms u^i_j as above, in 0.5, these axioms are satisfied. Conversely, if we start from a family (u^i_j) satisfying the conditions above, we shall see (in 3.5.9) that we can reconstruct the space $X = \mathrm{gl}\, U$ as the *gluing* of the diagram U, a quotient of the disjoint union of all U_i modulo the equivalence relation produced by the transition maps. (More precisely, the pair $(X, (u^i\colon U_i \to X))$ will be the *lax colimit* of the diagram (u^i_j), in the ordered category $\mathcal{C}$.)

The diagram U will often be written as $((U_i), (u^i_j))_I$, or as U_I. The family (u^i_j) is called the *intrinsic atlas*, or the *gluing atlas*, of the manifold.

In the ordered categories $\mathcal{C}$ and $\mathcal{C}^r$ (and in all the others used in this analysis of local structures), a prominent role will be played by the endomorphisms $e\colon X \to X$ which are restriction of identities, called *projectors*.

The projectors of X are idempotent endomorphisms and commute, forming a semilattice (i.e. an ordered set with all meets)

$$\mathsf{Prj}\,(X) = \{e\colon X \to\!\!\!\!\!\!\!+\;\, X \mid e \leqslant \mathrm{id}\, X\}, \qquad e \wedge e' = ee' = e'e. \tag{0.6}$$

In fact, the projectors determine the order: for parallel morphisms $f, g\colon X \nrightarrow Y$, the relation $f \leqslant g$, is equivalent to the existence of $e \in \mathsf{Prj}\,(X)$ such that $f = ge$. We can always take as e the *support* $\underline{e}(f)$ of f, namely the partial identity on Def f, or equivalently the least $e \in \mathsf{Prj}\,(X)$ such that $fe = f$.

These projectors satisfy axioms (see 3.3.5), and supply the categories $\mathcal{C}$ and $\mathcal{C}^r$ with the structure of an *e-cohesive category*, or *e-category*, one of the main ingredients of our analysis.

More precisely, $\mathcal{C}$ and $\mathcal{C}^r$ are *totally cohesive e-categories*, which means that every family of 'compatible' morphisms $f_i\colon X \nrightarrow Y$ has a join

$$\bigvee f_i\colon X \nrightarrow Y,$$

and composition distributes over these joins.

Being a compatible family can be simply read as 'upper bounded', but the important fact is that this property is characterised by supports. Namely, for $f, g\colon X \nrightarrow Y$, we say that f and g are *compatible*, or *linked* (written as $f\,!\,g$), if

$$f\,\underline{e}(g) = g\,\underline{e}(f), \tag{0.7}$$

which means that they coincide wherever they are both defined. If the morphisms $(f_i)_{i \in I}$ are pairwise linked, the join $f = \bigvee f_i$ is defined on $\bigcup \mathrm{Def}\, f_i$, and its graph is the union of the graphs of all f_i.

A general presentation of cohesive structures $(\leqslant, !)$ on categories can be found in Section 3.1, either determined by projectors or more general.

0.3 Morphisms of manifolds as linked profunctors

We can now define the category Mf $\mathcal{C}^r$, of C^r-manifolds and 'linked profunctors' between them, extending the formula (0.3), where a morphism $f\colon U \nrightarrow V$ is determined by its components

$$f_h^i = v_h f u^i\colon U_i \nrightarrow V_h,$$

on the charts of domain and codomain. A morphism in Mf $\mathcal{C}^r$ will be a 'linked profunctor', that is an enriched profunctor between enriched categories, satisfying a compatibility condition.

More precisely, an (enriched) *profunctor*

$$a = (a^i_h)_{IH} \colon (U_i, u^i_j)_I \to (V_h, v^h_k)_H \tag{0.8}$$

is a family of morphisms $a^i_h \colon U_i \to\!\!\!\!\!\!\bullet\; V_h$ in $\mathcal{C}^r$ such that, for all $i, j \in I$ and $h, k \in H$

(i) $a^j_h\, u^i_j \leqslant a^i_h, \quad v^h_k\, a^i_h \leqslant a^i_k$ (*profunctor laws*).

It will be said to be *linked*, or *compatible*, if it has a *resolution* $e_{ih} \in \mathsf{Prj}\, U_i$ $(i \in I,\ h \in H)$, defined by the property:

(ii) $a^i_k\, e_{ih} = v^h_k\, a^i_h$ (*left linking law*),

which is meant to ensure:

- that linked profunctors can be composed,
- that the gluing of a linked profunctor gives a *single-valued* partial mapping $a \colon \mathrm{gl}\, U \to\!\!\!\!\!\!\bullet\; \mathrm{gl}\, V$.

The resolution can be expressed by supports, taking $e_{ih} = \underline{e}(a^i_h)$.

The usual matrix composition of profunctors works, because $\mathcal{C}^r$ is totally cohesive: composing $a \colon U \to V$ with a consecutive linked profunctor $b \colon (V_h, v^h_k)_H \to (W_m, v^m_n)_M$, the composites

$$b^h_m\, a^i_h \colon U_i \to\!\!\!\!\!\!\bullet\; V_h \to\!\!\!\!\!\!\bullet\; W_m \qquad (h \in H),$$

form a linked family (for every $i \in I$, $m \in M$), and the component c^i_m of $c = ba \colon U \to W$ is computed as their linked join

$$c^i_m \colon U_i \to\!\!\!\!\!\!\bullet\; W_m, \qquad c^i_m = \textstyle\bigvee_{h \in H} b^h_m\, a^i_h. \tag{0.9}$$

This composition is based on resolutions (or supports): we cannot simply work in an ordered category with joins of upper bounded families of parallel morphisms — a sort of 'conditioned quantaloid'.

0.4 The interest of an intrinsic approach

This formalisation will allow us to move between different contexts.

For instance, the tangent bundle of an open n-dimensional euclidean space U is the trivial vector bundle $TU = U \times \mathbb{R}^n$. The present machinery automatically extends this obvious setting to the tangent functor of differentiable manifolds

$$T \colon \mathrm{Mf}\, \mathcal{C}^r \to \mathrm{Mf}\, \mathcal{V} \qquad (r > 0), \tag{0.10}$$

with values in the category of vector bundles, presented as intrinsic manifolds on an ordered category $\mathcal{V}$ of trivial vector bundles (in Section 4.2).

The same procedure works for tensor calculus.

In a more elementary way, the embedding of $\mathcal{C}^r$ in the category $\mathcal{C}$ of topological spaces and partial continuous mappings defined on open subspaces gives the topological realisation of C^r-manifolds

$$\mathrm{Mf}\,\mathcal{C}^r \to \mathcal{C} \tag{0.11}$$

taking into account that the second category is gluing complete (each manifold on $\mathcal{C}$ has a gluing space), and therefore $\mathcal{C}$ is equivalent to $\mathrm{Mf}\,\mathcal{C}$.

For a reader acquainted with the theory of enriched categories, we note that the property of Cauchy completeness of enriched categories, which is crucial in other contexts, is less important here where the morphisms are based on profunctors: replacing an intrinsic atlas by a complete one would simply give an isomorphic object (with respect to profunctors).

Furthermore, this can only be done when the basis of enrichment is a small category: it is the case for topological or differentiable manifolds, but not for fibre bundles. (Cauchy completion of enriched categories is reviewed in Chapter 6.)

0.5 An outline

Every chapter and every section has its own introduction; this is a brief synopsis, and involves topics which may be unknown to the reader.

Chapter 1 introduces the theory of ordered sets, semigroups and categories, as far as needed in this book. Some care is devoted to the classical theory of inverse semigroups. Limits and colimits in categories are only examined in their basic forms; adjoint functors are briefly presented.

Chapter 2 is devoted to ordered categories, equipped with a local order between parallel morphisms, as in Section 0.1.

The main part of the chapter deals with *inverse categories* and their canonical order, a natural extension of inverse semigroups. In fact, our categories of partial mappings, like $\mathcal{C}$ and $\mathcal{C}^r$ (see 2.1.4), have an *inverse core*, $\mathrm{I}\mathcal{C}$ and $\mathrm{I}\mathcal{C}^r$, formed of the 'partial isomorphisms' of the category (see 2.4.4, 2.4.5).

The symmetry law 0.2(iii) forces the transition maps of a manifold to belong to the inverse core; but we need the whole categories $\mathcal{C}$ and $\mathcal{C}^r$ to construct the general morphisms of manifolds, as presented above, in Section 0.3.

Some topological prerequisites are also reviewed in this chapter.

Chapter 3 introduces and studies 'cohesive categories', as ordered categories equipped with a structure which allows us to build our categories of local structures, like $\mathrm{Mf}\,\mathcal{C}^r$.

We are mainly interested in e-cohesive categories, where the order $f \leqslant g$, and the linking relation $f\,!\,g$, of morphisms are determined by their supports $\underline{e}(f)$ and $\underline{e}(g)$. But we also give a more general, unifying notion of cohesive category, because the linking relation of the inverse cores $\mathrm{I}\mathcal{C}$ and $\mathrm{I}\mathcal{C}^r$ is not determined in this way: it also needs cosupports, on codomains.

Theoretically, the main results are the gluing completion theorems 3.5.8 and 3.6.7, which give the universal properties of the categories of manifolds built in this chapter.

Chapter 4 shows how various concrete local structures (and their interplay) can be formalised in this way: topological and differentiable manifolds, manifolds with boundary, foliated manifolds, fibre bundles, vector bundles, G-bundles, simplicial complexes, etc.

New developments, inspired by Directed Algebraic Topology [G8], deal with 'locally cartesian ordered manifolds', which are spaces with distinguished paths, generally non-reversible (in Sections 4.3–4.5).

Chapter 5 gives further information on category theory. On this basis, Chapter 6 studies the relationship of our approach to manifolds with the general theory of enriched categories.

Finally, Chapter 7 collects the solutions of most exercises.

0.6 Manifolds by Ehresmann's pseudogroups

Loosely speaking, the approach of C. Ehresmann to differentiable manifold and other local structures relies on categories of *total mappings*, equipped with a *global order* (examined in Sections 2.8 and 6.7).

(a) First we need the groupoid $\mathrm{Iso}(\mathsf{Top}) = \mathrm{Iso}(\mathcal{C})$ of *topological spaces and homeomorphisms*, equipped with a 'global order' $f' \subset f$ on its maps, defined as follows: the homeomorphism $f' \colon X' \to Y'$ is the restriction of the homeomorphism $f \colon X \to Y$, from an open subspace X' of X to an open subspace $Y' = f(X')$ of Y.

(b) Similarly we have the groupoid $\mathrm{Iso}(\mathcal{C}^r) = \mathrm{Iso}(\mathcal{C}) \cap \mathcal{C}^r$ of *open euclidean spaces and* C^r*-diffeomorphisms*, with the restricted global order.

A *chart* is again a total homeomorphism in $\mathrm{Iso}(\mathcal{C})$

$$u^i \colon U_i \to X_i, \qquad (0.12)$$

as in the classical definition recalled above, with inverse $u_i \colon X_i \to U_i$.

To make sense of the condition expressed in (0.2) we resort to the *extended composition*, or *pseudoproduct*

$$u_j{\bullet}u^i\colon u_i(X_i \cap X_j) \to u_j(X_i \cap X_j), \tag{0.13}$$

namely the homeomorphism that takes $u_i(x)$ to $u_j(x)$, for all $x \in X_i \cap X_j$.

This extended composition turns the set of morphisms of $\mathrm{Iso}(\mathcal{C})$ into an (inverse) semigroup; one can then require the composite to belong to $\mathrm{Iso}(\mathcal{C}^r)$.

One constructs in this way a groupoid of manifolds and diffeomorphisms of class C^r.

This approach makes a deep use of formal set theory, which we prefer to avoid: dealing with maps $f\colon X \to Y$, $g\colon Z \to W$ between arbitrary spaces the meaning of the relation $f \subset g$ and of the pseudoproduct $f{\bullet}g$ seems to be unclear, unless all these spaces are known to be subspaces of a given space.

This can be managed for Hausdorff paracompact differentiable manifolds, using the Whitney embedding theorem, according to which an n-dimensional manifold of this kind can always be embedded in $\mathbb{R}^{2n}$ (as exploited in Section 4.6). In other cases, for instance for fibre bundles, there is no opportunity of this kind.

0.7 Prerequisites, notation and conventions

This book is addressed to readers with different formation, in Topology, or Differential Geometry, or Category Theory, or Semigroup Theory, or perhaps other fields; at the cost of dealing with aspects that can be obvious to one reader or another.

We only assume as known the basic theory of topological spaces, differentiable manifolds, abelian groups, modules and vector spaces. Ordered sets, semigroups and categories are introduced and studied as far as needed here. Banach spaces occur in a marginal way. Deeper results, when used, are referred to.

The symbol $\subset$ denotes *weak* inclusion. A singleton set is often written as $\{*\}$. The equivalence class of an element x, with respect to an assigned equivalence relation, is generally written as $[x]$. A bullet in a diagram stands for an object.

We write as $|X|$ the underlying set of a structured set X, e.g. a topological space, or a semigroup. Dealing with topological spaces, the term *map* will often be used for 'continuous mapping'; neighbourhood can be abridged to 'nbd'.

A ring R is assumed to be unital, and a (left) R-module is assumed to be unitary.

The symbols $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ denote the sets of natural, integral, rational, real or complex numbers. The topology of the standard euclidean sphere $\mathbb{S}^n$ is reviewed in Section 2.5.

The standard compact interval $[0, 1]$ is also written as $\mathbb{I}$. Open and semi-open real intervals are denoted as $]a, b[$, $[a, b[$, etc. — a notation, which distinguishes the open interval $]a, b[$ from the pair (a, b), as in Bourbaki's treatise.

Categories of partial mappings, like $\mathcal{S}$ and $\mathcal{C}$, play a central role (see Section 2.1). Our analysis of these categories will be based (as in Section 0.3) on particular idempotent endomorphisms $e\colon X \nrightarrow X$, the 'partial identities', called *projectors*. Of course these morphisms should not be confused with the projections of a cartesian product, nor with the projection on a quotient.

A part marked with * is out of the main line of exposition. It may refer to issues dealt with in the sequel, or be addressed to readers with some knowledge of the subject, or give references to higher topics.

Most exercises have a solution or convenient hints. These can be found in Chapter 7, or — occasionally — below the exercise, if they are important for the sequel. Easy exercises may be left to the reader.

0.8 Sources and outgrowth

Our presentation of manifolds as intrinsic atlases in cohesive categories, in Chapters 3 and 4, is an expansion of matter published in two articles, in 1989–90 [G3, G4], partially based on a long work on inverse categories in Homological Algebra, in the 1970's and 1980's. (The results of the latter are summarised in [G2], and exposed in the recent book [G9].)

Presenting manifolds in an intrinsic way can be found in the literature. But the roots of our approach rely on two main domains.

(i) A first source was Ehresmann's work on local structures, in the 1960's.

(ii) Another main source is the theory of enriched categories [EiK, Kl2], and Lawvere's claim that many interesting mathematical structures (besides forming categories) *are* themselves categories, enriched on a suitable basis: a *monoidal* category as in Lawvere's article, in 1974 [Lw], or more generally a *bicategory* as in many subsequent papers [Bet, Wa1, Wa2, BetC, BetW1, BetW2].

The bases we actually use are very particular bicategories: ordered categories with 'linked' joins, preserved by composition.

Other papers of the 1980's are related to our approach, or to the structures we are using.

(a) S. Kasangian and R.F.C. Walters worked in the perspective opened by Lawvere, aiming to present differentiable manifolds as symmetric enriched categories on an involutive ordered category with all joins of parallel map, and to explore Cauchy-completeness in this context. Their research was presented in an (unpublished) talk at a Surrey meeting on Category Theory, in 1982 [KaW].

Constructing categories similar to $\mathcal{C}$ and $\mathcal{C}^r$, but having all joins (rather than the linked ones), leads to complications. Something of this kind will be presented in Section 4.6, taking advantage of the fact that n-dimensional differentiable manifolds can be embedded in $\mathbb{R}^{2n}$, where the transition maps can be treated as partial identities, and have arbitrary joins. General local structures cannot be dealt with in this way.

(b) Dominical categories and p-categories, other formalisations of categories of partial mappings, were introduced in the 1980's by R.A. Di Paola, A. Heller, G. Rosolini and E. Robinson [Di, He, DiH, Rs, RoR], making use of a monoidal structure derived from cartesian products. These categories have a natural e-structure (see 3.3.9(c)).

Finally, after the introduction of e-categories in [G3, G4], this structure has been used in computer science and category theory, under the name of 'restriction category' and equivalent axioms [CoL]: see 3.3.9(d). Later, totally cohesive e-categories have also been used, under the name of 'join-restriction categories'. A recent paper acknowledges the fact that restriction categories are the same as e-categories [CoG].

0.9 Acknowledgements

As already said, this book, is mainly indebted to Ehresmann's approach to local structures, and Lawvere's unusual view of enriched categories (partially exposed in Chapter 6).

When I presented this approach to 'manifolds', at the 1988 Prague Conference on 'Categorical Topology', Mac Lane rightly objected that the provisional term I was using, 'coherent category', might lead to confusion with categorical coherence theorems. Soon after I found a replacement: 'cohesive' instead of 'coherent', already adopted in the Proceedings of this conference [G3].

I remember with pleasure a letter by Andrée C. Ehresmann, which prompted me to sketch in [G4] a comparison between e-categories and C. Ehresmann's categories with a 'global order'. This comparison is now developed in Section 6.7.

I am pleased to acknowledge several helpful discussions with my colleague Ettore Carletti.

I would also like to thank Dr Lim Swee Cheng and Ms Soh Jing Wen, at World Scientific Publishing Co., for their kind, effective help in the publication of this book. The cover is due to Mr Ng Chin Choon.

Diagrams and figures are composed with 'xy-pic', a free package by K.H. Rose and R. Moore.

1
Order, semigroups and categories

This is an introductory chapter on ordered sets, semigroups, inverse semigroups and categories. Many points will be obvious to one reader or another, according to their interests.

The choice of arguments is aimed at the present applications, and by no means representative of Order Theory, or Semigroup Theory, or Category Theory. Complements on categories will be added in Chapters 5 and 6, related to topics that appear in the previous parts.

We recall that the symbol $\subset$ always denotes weak inclusion. The basic theory of topological spaces, groups, rings and modules is assumed to be known. Solutions of non-obvious exercises can generally be found in the last chapter. A part marked with * refers to developments which are not technically needed, or are referred to.

1.1 Preordered sets, lattices and semigroups

We begin by reviewing the basic notions of preordered sets and semigroups; we also examine the interplay of preorders and topology.

Further information on lattice theory can be found in Birkhoff [Bi] and Grätzer [Gr]; on semigroup theory in Clifford–Preston [CP], Howie [Ho] and Lawson [Ls].

1.1.1 Preordered and ordered sets

We use the following terminology.

A *preordered set* X is a set equipped with a *preorder relation* $x \prec x'$ (read as x *precedes* x'), which is assumed to be reflexive and transitive. It is an *ordered set* if the relation is *anti-symmetric*: if $x \prec x'$ and $x' \prec x$, then $x = x'$; an order relation is more often written as $x \leqslant x'$. If useful, one can write $x \prec_X x'$ and $x \leqslant_X x'$. A symmetric preorder relation is an equivalence relation, often written as $x \sim x'$.

In a *totally* ordered set any two elements are comparable: $x \leqslant x'$ or $x' \leqslant x$. An ordered set is often called a '*partially* ordered set', abbreviated to 'poset', to mean that totality is not assumed (but not excluded).

Every set X has a *discrete order* $x = x'$, which is the *finest*, or least preorder relation. It also has an *indiscrete*, or *chaotic preorder*, the relation $x, x' \in X$, which is the *coarsest*, or greatest preorder relation on X.

A preordered set X has an associated equivalence relation $x \sim x'$ defined by the conjunction: $x \prec x'$ and $x' \prec x$. The quotient $X/{\sim}$ has an induced order:

$$[x] \leqslant [x'] \quad \Leftrightarrow \quad x \prec x'. \tag{1.1}$$

If X is a preordered set, X^{op} is the *opposite*, or *dual* one — with reversed preorder. Every topic of the theory of preordered sets has a dual instance, which comes from the opposite preordered set, or sets.

Let X be a preordered set. The *minimum* $\min X$ is an element which precedes all the elements of X (and can exist or not, of course); the *maximum* $\max X$ is an element preceded by all the elements of X. They are determined up to the associated equivalence relation in X, and uniquely determined if X is ordered. They can also be written as $\bot$ and $\top$ (*bottom* and *top*).

Every subset of a preordered set will be equipped with the restricted preorder, by default.

A mapping $f \colon X \to Y$ between preordered sets is said to be *monotone*, or *preorder-preserving*, or (weakly) *increasing*, if $x \prec_X x'$ implies $f(x) \prec_Y f(x')$, for all $x, x' \in X$. It is *isotone* if it preserves and reflects the preorder: $x \prec_X x'$ if and only if $f(x) \prec_Y f(x')$. (If X is an ordered set, this implies that f is injective.)

An *isomorphism* $f \colon X \to Y$ of preordered sets is a bijective monotone mapping whose inverse mapping $f^{-1} \colon Y \to X$ is also monotone; equivalently, f is an isotone bijection. More generally, an *embedding* $f \colon X \to Y$ of preordered sets is an injective isotone mapping, and gives an isomorphism from X to the preordered subset $f(X) \subset Y$.

Examples (a) The set $\mathbb{R}$ of real numbers, equipped with the natural order $x \leqslant y$, is called the *ordered line*. It is a totally ordered set. Its cartesian power $\mathbb{R}^n$ has a canonical (partial) order:

$$(x_1, ..., x_n) \leqslant (y_1, ..., y_n) \;\; \Leftrightarrow \;\; (\text{for all } i = 1, ..., n\colon\; x_i \leqslant y_i), \tag{1.2}$$

that will be called the *cartesian order* of $\mathbb{R}^n$.

(b) More generally, any cartesian product $X = \prod_i X_i$ of preordered sets has a canonical preorder, defined componentwise as above: $(x_i) \prec_X (y_i)$ if and only if, for all indices i, $x_i \prec y_i$ in X_i. The preorder of X is the coarsest that makes all projections $X \to X_i$ monotone.

(c) For every set S, the power set $\mathcal{P}S$ is ordered by inclusion, with minimum $\emptyset$ and maximum S. This order is not total, as soon as S has more than 1 element.

(d) In the set $\mathbb{N}$ of natural numbers, the divisibility relation $n \mid m$ means that $m = nn'$ for some $n' \in \mathbb{N}$; it is an order relation, with minimum 1 and maximum 0. In the set $\mathbb{Z}$ of integers the divisibility relation is a preorder; the associated equivalence relation is $k = \pm k'$, and the associated ordered set $\mathbb{Z}/{\sim}$ is isomorphic to $\mathbb{N}$ (ordered by divisibility). The minimum of $\mathbb{Z}$ is 1 (or -1), the (unique) maximum is 0.

1.1.2 Infima and suprema

Let X be an ordered set. For $A \subset X$ and $a \in X$, the sets of their *lower bounds* and their *upper bounds* in X will be written as

$$\begin{aligned} L(A) &= \{x \in X \mid x \prec a, \text{for all } a \in A\}, \\ {\downarrow}a &= L(\{a\}) = \{x \in X \mid x \prec a\}, \\ U(A) &= \{x \in X \mid a \prec x, \text{for all } a \in A\}, \\ {\uparrow}a &= U(\{a\}) = \{x \in X \mid a \prec x\}. \end{aligned} \qquad (1.3)$$

The *infimum* $\inf_X A$ of A in X, or *greatest lower bound*, or *meet*, is defined as

$$\inf_X A = \max\,(L(A)), \qquad (1.4)$$

also written as $\inf A$ or $\bigwedge A$. Dually, the *supremum* $\sup_X A$ of A in X, or *least upper bound*, or *join*, is

$$\sup_X A = \min\,(U(A)) = \inf_{X^{\mathrm{op}}} A, \qquad (1.5)$$

also written as $\sup A$ or $\bigvee A$. (These outcomes can exist or not.) If A has a minimum m, the infimum also exists and $\inf_X A = \max\,({\downarrow}m) = \min A$; if A has a maximum, $\sup_X A = \max A$.

Every element of X is (trivially) a lower bound and an upper bound of the empty subset; if X has a greatest and a least element, we have:

$$\inf_X \emptyset = \max X = \sup_X X, \qquad \sup_X \emptyset = \min X = \inf_X X. \qquad (1.6)$$

These definitions can be extended to preordered sets, but then joins and meets are only determined up to the equivalence relation associated to our preorder.

A *meet semilattice*, or *lower semilattice*, X is an ordered set where every pair $x, y \in X$ has a *meet* $x \wedge y = \inf\{x, y\}$; a *1-semilattice* is also assumed to have a top element, written as 1 (or $\top$), which acts as a unit for the meet operation. Equivalently, a 1-semilattice is an ordered set where every finite subset A has a greatest lower bound $\bigwedge A$.

Dually, a *join semilattice*, or *upper semilattice*, has all binary joins $x \vee y$; a *0-semilattice* is also assumed to have a bottom element, written as 0 or $\bot$; in other words, a 0-semilattice is an ordered set where every finite subset A has a least upper bound $\bigvee A$.

An ordered set is said to be *filtered* (or *directed*) if every pair $x, y \in X$ has an upper bound; dually, it is *cofiltered* if every pair has a lower bound.

1.1.3 From lattices to boolean algebras

A *lattice* is an ordered set where every pair $x, y \in X$ has a *meet* $x \wedge y = \inf\{x, y\}$ and a *join* $x \vee y = \sup\{x, y\}$.

A *bounded lattice* is also required to have a top element, written as 1 (or $\top$), and a bottom element 0 (or $\bot$), that are units for the meet and join operation, respectively. (In category theory the term 'lattice' is often used in this sense; we do not follow this convention here.)

A lattice is said to be *distributive* if the meet operation distributes over the join operation

(D) $\quad (x \vee y) \wedge z = (x \wedge z) \vee (y \wedge z),$

or equivalently if the join distributes over the meet (see Exercise 1.1.4(a)).

A *boolean algebra* is a distributive bounded lattice where every element x has a *complement* x', defined by the following properties

$$x \wedge x' = 0, \qquad x \vee x' = 1, \tag{1.7}$$

and determined by them (as proved in Exercise 1.1.4(b)).

A *complete lattice* X is an ordered set where every subset has a join and a meet. But actually it is sufficient to assume the existence of all meets (or all joins), since one can recover the join of a subset $A \subset X$ as the meet of its upper bounds, or symmetrically (as proved in Exercise 1.1.4(d))

$$\bigvee A = \bigwedge(U(A)), \qquad \bigwedge A = \bigvee(L(A)). \tag{1.8}$$

More precisely, the first formula above means that, for any subset A of an ordered set X, $\bigvee A$ exists if and only if $\bigwedge(U(A))$ exists, and — in this case — they coincide.

1.1.4 Exercises and complements

The non-obvious solutions of these exercises can be found in Chapter 7.

(a) In a lattice X, the property 1.1.3(D) implies that the join operation distributes over the meet.

(b) In a distributive bounded lattice X, the complement of x, defined in (1.7), is unique.

(c) In an ordered set X, the lower and upper bounds of subsets $A, B \subset X$ satisfy these properties

$$\begin{gathered} A \subset B \quad \Rightarrow \quad L(A) \supset L(B) \text{ and } U(A) \supset U(B), \\ A \subset UL(A), \qquad A \subset LU(A), \\ LUL(A) = L(A), \qquad ULU(A) = U(A). \end{gathered} \tag{1.9}$$

The mappings $L, U \colon \mathcal{P}X \to \mathcal{P}X$ form thus a contravariant Galois connection, see 1.8.8(e).

(d) Prove the formulas (1.8).

(e) A totally ordered set X is always a lattice, with $x \wedge y = \min\{x, y\}$ and $x \vee y = \max\{x, y\}$; it is a distributive lattice.

(f) The ordered line $\mathbb{R}$ and its (partially ordered) cartesian power $\mathbb{R}^n$ are distributive lattices, without minimum and maximum. They are *conditionally complete*: every non-empty upper bounded subset has a join, and every non-empty lower bounded subset has a meet. Prove that these two properties are equivalent, in every ordered set X.

(g) Prove that an ordered set X has all joins of upper bounded subsets if and only if it has all meets of non-empty subsets. The interval $[0, +\infty[$ of the ordered line has this form of 'conditioned completeness', and lacks a maximum.

(h) In a power set $\mathcal{P}S$, a subset $\mathcal{A} \subset \mathcal{PS}$ is more easily understood as an indexed family $(A_i)_{i \in I}$ of subsets of S. A lower bound of this family is any subset contained in all of them, and the greatest is $\bigcap A_i$. Symmetrically, the least upper bound of the family is $\bigcup A_i$.

$\mathcal{P}S$ is a complete boolean algebra, where joins and meets are unions and intersections, while the boolean complement of a subset A is the set-theoretical complement $X \setminus A$.

(i) The ordered set $\mathrm{Sub}(A)$ of submodules of an R-module A has also all meets, which are intersections: $\bigwedge A_i = \bigcap A_i$. Therefore $\mathrm{Sub}(A)$ is a complete lattice, and $\bigvee A_i$ is the least submodule of A containing the set-theoretical union $\bigcup A_i$.

The join is often written as $\sum A_i$, because it can be realised as the set of elements of A which are sums $x_1 + x_2 + ... + x_n$ of elements of $\bigcup A_i$.

(j) Show that the lattice $\mathrm{Sub}(A)$ of subgroups of an abelian group need not be distributive. (It has a weaker distributive property, dealt with in 1.1.9.) *Hints:* take $A = \mathbb{Z}^2$.

*(k) Frames and quantales will be treated in Section 6.1.

1.1.5 Semigroups

A *semigroup* S is a set equipped with an associative operation, generally written in multiplicative form, as ab or $a.b$. We do *not* exclude the *empty semigroup*, as is often the case in semigroup theory (cf. [CP], p. 1) and universal algebra.

If S is a semigroup, S^{op} is the *opposite* or *dual* one, with reversed product $a * b = ba$. As for preorders, every topic or statement of the theory of semigroups has a dual instance.

For two subsets $A, B \subset S$ and an element $x \in S$, one writes

$$\begin{gathered} AB = \{ab \mid a \in A,\ b \in B\} \subset S, \\ xA = \{x\}A = \{xa \mid a \in A\}, \qquad Ax = A\{x\} = \{ax \mid a \in A\}. \end{gathered} \tag{1.10}$$

In a semigroup S, a *subsemigroup* is a subset T closed under product: $TT \subset T$. The intersection $\bigcap S_i$ of a family of subsemigroups of S is a subsemigroup (also because we are not excluding the empty subsemigroup). Therefore, the set $\mathrm{Sub}(S)$ of subsemigroups of S forms a complete lattice with respect to inclusion; its minimum is the empty subsemigroup.

If A is a subset of S, the *subsemigroup* $\langle A \rangle$ *of* S *generated by* A is the least subsemigroup of S that contains A. Formally, it is the intersection of all the subsemigroups of S that contain A; concretely, it consists of all products $x_1 x_2 ... x_n$ of elements of A (for n a positive integer). The join $\bigvee S_i$ of a family of subsemigroups of S is generated by their union $\bigcup S_i$.

A *homomorphism* $h \colon S \to T$ of semigroups is a mapping that preserves the product: $h(ab) = h(a).h(b)$, for all $a, b \in S$. Its image $h(S)$ is a subsemigroup of T.

1.1.6 Unit and absorbing element

The semigroup S is called a *unital semigroup*, or a *monoid*, when it has a unit 1 (acting as an identity for the product), that is obviously unique.

The unit is also written as 1_S when useful. A *homomorphism of monoids* is assumed to preserve the unit.

Every semigroup S has an *associated monoid* S^1, obtained by adding an element $1 \notin S$ and extending the product in the obvious way: $1a = a = a1$ for all $a \in S^1$. Note that if S already has a unit e, the latter is no longer a unit in S^1.

The *universal property* of this construction says that every semigroup-homomorphism $f\colon S \to M$ with values in a monoid can be uniquely extended to a monoid-homomorphism $g\colon S^1 \to M$, letting $g(1) = 1_M$. *(Universal properties are formalised in 1.5.7.)*

(Our convention on this point differs from the usual one in semigroup theory, where one takes $S^1 = S$ when S is already unital. In this way the previous universal property is not satisfied, and the procedure $S \mapsto S^1$ cannot be extended to homomorphisms.)

An *absorbing element* for the semigroup S is an element z such that $za = z = az$ for all $a \in S$. This uniquely determined element can be written as 0 or ∞, as convenient.

Again, every semigroup S has an *associated semigroup with an absorbing element* S^∞, obtained by adding an element $\infty \notin S$ and extending the product in the obvious way: $\infty a = \infty = a\infty$ for all $a \in S^\infty$. There is again an obvious universal property.

1.1.7 Idempotents

In a semigroup S the *idempotent* elements e (such that $ee = e$) play an important role, and should be viewed as 'partial identities': see 1.2.4. *(This role will be formalised by a category, the 'idempotent completion' of S, in 2.2.5.)*

The set of idempotent elements of S, often denoted by E (or E_S) has a *canonical order* $e \leqslant f$, characterised by the following conditions, which are easily seen to be equivalent:

(i) $e = ef = fe$,

(ii) $e = fef$,

(iii) $e \in Sf \cap fS$,

(iv) $e \in fSf$.

The set E is not closed under product in S, generally. But the product of two *commuting* idempotents e, f is always idempotent

$$ef = fe \;\; \Rightarrow \;\; ef.ef = ee.ff = ef, \tag{1.11}$$

and then the element $ef = fe$ is the meet $e \wedge f$, with respect to the canonical order of E.

If S is a monoid, the unit 1 is the greatest element of E_S. An absorbing element is the least element of E_S.

Every homomorphism $h \colon S \to T$ of semigroups restricts to a mapping $E_S \to E_T$ that preserves the canonical order.

An *idempotent semigroup* (also called a *band* in semigroup theory) is a semigroup where each element is idempotent.

1.1.8 *Preorder and Alexandrov topologies

The reader may have noticed a similarity of the theories of preordered sets and topological spaces. In fact, the former can be 'embedded' in the latter, so that — for instance — the finest preorder corresponds to the finest topology, the coarsest to the coarsest.

Let X be an arbitrary topological space. For $a \in X$, we write as $\overline{a}$ the closure of the singleton $\{a\}$, also called the closure of the point a. The *specialisation preorder* of the space X is defined by

$$x \prec y \quad \Leftrightarrow \quad x \in \overline{y} \quad \Leftrightarrow \quad \overline{x} \subset \overline{y}, \tag{1.12}$$

so that $\overline{a} = \{x \in X \mid x \prec a\}$.

Generally speaking, this preorder misses a large part of the information contained in the topology: for instance, in a space where all singletons are closed, this preorder is discrete: $x = y$. Thus, on the set $\mathbb{R}$, the euclidean and the discrete topology give the same order relation.

This is no longer the case if we restrict in a convenient way the topologies we are considering.

An *Alexandrov topology*, named after Pavel S. Alexandrov, is a topology where the open sets are stable under arbitrary unions *and arbitrary intersections*; equivalently, the same is true of closed sets.

In this case, for any subset $A \subset X$

$$\overline{A} = \textstyle\bigcup_{a \in A} \overline{a}, \tag{1.13}$$

because this union is closed, and obviously the least closed subset containing A. Therefore, an Alexandrov topology is determined by the closure of its points, and also by the specialisation preorder. The topology is T_0 if and only if the specialisation preorder is an order relation.

The reader may now guess that preordered sets are 'the same as' spaces with Alexandrov topology, and monotone mappings amount to continuous mappings between such spaces.

More precisely, the previous argument makes the category of preordered sets isomorphic to a full subcategory of the category of topological spaces, formed of the spaces with Alexandrov topology. Similarly, ordered sets correspond to spaces with an Alexandrov T_0-topology.

All this is well known; details can be found in [G11], 3.2.7 and 6.5.4.

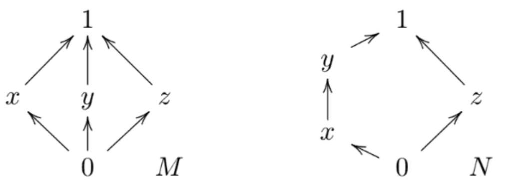
*1.1.9 *Modular lattices*

The lattice Sub(A) of submodules of an R-module is not distributive, generally; but one can easily check that it always satisfies a weaker, restricted form of distributivity, called 'modularity' (as proved in Exercise (a)).

Namely, a lattice is said to be *modular* if it satisfies the following selfdual property (for all elements x, y, z)

(M) if $x \leqslant z$ then $(x \vee y) \wedge z = x \vee (y \wedge z)$.

For instance, if A is the Klein four-element group $\mathbb{Z}/2 \oplus \mathbb{Z}/2$, the lattice Sub$(A)$ is isomorphic to the lattice $M = \{0, x, y, z, 1\}$ of five elements displayed in the following diagram, at the left



(1.14)

where the relation $a \leqslant b$, is represented by an arrow $a \to b$, and the meet (resp. join) of any two elements among x, y, z is 0 (resp. 1). This lattice is not distributive: note that x has two distinct complements, y and z. On the other hand, the lattice N displayed above is not even modular, as $x \leqslant y$ but $(x \vee y) \wedge z = z$ and $x \vee (y \wedge z) = x$.

Exercises and complements. (a) Every lattice Sub(A) of submodules is modular.

*(b) In figure (1.14), N is the 'basic' non-modular lattice, in the sense that: a lattice is modular if and only if it does not contain a sublattice isomorphic to N ([Bi], I.7, Theorem 12).

Similarly, M is the basic non-distributive *modular* lattice: a modular lattice is distributive if and only if it does not contain a sublattice isomorphic to M ([Bi], II.8, Theorem 13).

These two theorems can be combined, to characterise distributive lattices among all lattices.

1.2 Regular and inverse semigroups

After introducing *regular* and *inverse semigroups*, we prove a classical result: inverse semigroups are the same as regular semigroups with commuting idempotents (Vagner–Liber Theorem, in 1.2.6).

Then we briefly introduce *orthodox semigroups*, a more general notion.

The theory of orthodox, inverse and quasi-inverse semigroups was developed by B.M. Schein [Sc], N.R. Reilly and M.E. Scheiblich [ReS], T.E. Hall [Ha], M. Yamada [Ya], A.H. Clifford and G.B. Preston [CP], and others, starting in the 1960's. It can now be found in Howie's and Lawson's books [Ho, Ls].

1.2.1 Regular semigroups

One says that the elements a, b of the semigroup S are *partial inverses*, or form a *regular pair*, if

$$a = aba, \qquad\qquad b = bab. \tag{1.15}$$

The semigroup S is said to be *regular* if every element has some partial inverse (which need not be unique). This property, introduced by von Neumann [vN] in 1936 for the multiplicative semigroup of a ring, will also be called *von Neumann regularity.*

Note that each of the two relations above, in (1.15), implies that *ab and ba are idempotent elements of S*. On the other hand, every idempotent element e is a partial inverse of itself; therefore, an idempotent semigroup is always regular.

A homomorphism of semigroups preserves regular pairs.

We shall see, in 1.3.3, that the endo-relations of any R-module form a regular monoid.

1.2.2 Proposition (Regular semigroups)

(a) The semigroup S is regular if (and only if) for every $a \in S$ there is some $b \in S$ such that $a = aba$.

(b) Let S be a regular semigroup. If the set E_S has a maximum 1, this is a unit for S.

Note. The property in (a) is often taken as the definition of a regular semigroup.

Proof (a) If this property holds, then ab and ba are idempotents.

It follows that $b' = bab$ is a partial inverse of a:

$$ab'a = a(bab)a = (ab)(ab)a = aba = a,$$
$$b'ab' = (bab)a(bab) = b(ab)(ab)(ab) = bab = b'.$$

(b) If $1 = \max E_S$, then $e = e1 = 1e$ for every idempotent e. Any element a can be written as $a = a(ba) = (ab)a$ where ba and ab are idempotents, so that $a = a1 = 1a$. □

1.2.3 Inverse semigroups

The semigroup S is said to be *inverse* if every element a has precisely one partial inverse, which we write as $a^\sharp$. (In semigroup theory $a^\sharp$ is usually said to be 'inverse' to a; we keep the term 'partial inverse'.)

In this case, an element a is said to be *symmetric* if $a = a^\sharp$, which is equivalent to $a = a^3$. In particular, every idempotent e is a symmetric element: $e = e^\sharp$, but there can be symmetric elements which are not idempotent (see below). Any element $a \in S$ has two associated idempotents, $a^\sharp a$ and $aa^\sharp$, which will be further analysed below (in 1.3.2); moreover, a is idempotent if and only if $a = a^\sharp a$, if and only if $a = aa^\sharp$.

By definition, an *inverse subsemigroup* of an inverse semigroup S is a subsemigroup closed under partial inverses. A homomorphism between inverse semigroups automatically preserves the partial inverses (and is thus a homomorphism of inverse semigroups).

Exercises and complements. (a) If S is a monoid, an invertible element a has a unique partial inverse: its inverse a^{-1}.

(b) Every group is an inverse semigroup; more precisely, a group is the same as an inverse monoid were the left (or the right) cancellation law holds.

(c) In a group, the symmetric elements are the involutive ones, with $a^2 = 1$, and there is only one idempotent, the unit.

(d) The monoid $S(X)$ of endomappings of any set X is a regular semigroup.

(e) The monoid $S(X)$ is not inverse, except trivial cases (to be determined).

1.2.4 Partial bijections of a set

We show now that every set X has an associated inverse semigroup $\mathcal{I}(X)$ of 'partial bijections', introduced by V.V. Vagner (also transliterated as Wagner) in 1952 [Va1], and called the *symmetric inverse semigroup* of X, by analogy with the symmetric group of permutations of X.

In analogy with Cayley theorem for groups, we shall see that every inverse semigroup S can be embedded as an inverse subsemigroup of $\mathcal{I}(S)$, in 1.3.6.

A *partial bijection* of the set X will be a bijective mapping $f\colon U \to V$ between subsets U, V of X (possibly empty). We write

$$\mathrm{Def}\, f = U, \qquad \mathrm{Val}\, f = V. \tag{1.16}$$

(Equivalently, it is an injective mapping $U \to X$ defined on a subset of X; but we shall not use the term 'partial injection', which can mask the symmetric character of the notion.)

These mappings form the set $\mathcal{I}(X)$. Given another bijective mapping $g\colon V' \to W$ between subsets of X, the *composite* gf is again a partial bijection of X, defined where it makes sense

$$\mathrm{Def}\,(gf) = \{x \in X \mid x \in U, f(x) \in V'\}, \qquad (gf)(x) = g(f(x)). \tag{1.17}$$

The operation is associative: $\mathcal{I}(X)$ is a monoid, with unit the identity mapping $\mathrm{id}\, X$ (everywhere defined, of course). The empty bijection 0 is the *absorbing element* of the semigroup: $0f = 0 = f0$, for every $f \in \mathcal{I}(X)$ (cf. 1.1.6).

Finally, $\mathcal{I}(X)$ is an inverse monoid: a bijective mapping $f\colon U \to V$ between subsets of X has a unique partial inverse in $\mathcal{I}(X)$, represented by the inverse mapping $f^\sharp\colon V \to U$. Let us note that the associated idempotents

$$f^\sharp f = \mathrm{id}\, U, \qquad ff^\sharp = \mathrm{id}\, V, \tag{1.18}$$

coincide with the unit $\mathrm{id}\, X$ if and only if f is a *total* bijection $X \to X$. Only in this case the partial inverse $f^\sharp$ is a true inverse, and will be written as f^{-1}.

We have already seen that an element $f \in \mathcal{I}(X)$ is idempotent if and only if $f = f^\sharp f$. Therefore the idempotents of the symmetric inverse semigroup $\mathcal{I}(X)$ coincide with the partial identities $\mathrm{id}\,(U)$ of X, i.e. the identities of the subsets $U \subset X$.

1.2.5 Other examples

If X is a topological space we are interested in the set $\mathrm{I}\mathcal{T}(X)$ of its *partial homeomorphisms*, i.e. homeomorphisms $f\colon U \to V$ between subspaces of X, and (even more) in the subset $\mathrm{I}\mathcal{C}(X)$ of homeomorphisms $f\colon U \to V$ between *open* subspaces of X.

The notation $\mathrm{I}\mathcal{T}$, $\mathrm{I}\mathcal{C}$ will be explained in Definition 2.4.4. *(It stands for the 'inverse core' of two categories, $\mathcal{T}$ and $\mathcal{C}$, of partial continuous mappings. Similarly, $\mathcal{I}$ is the inverse core of the category $\mathcal{S}$ of sets and partial mappings.)*

Writing the underlying set of the space X as $|X|$, we have the following inclusions of inverse subsemigroups:

$$\mathrm{I}\mathcal{C}(X) \subset \mathrm{I}\mathcal{T}(X) \subset \mathcal{I}(|X|).$$

1.2.6 Theorem (Vagner–Liber Therem on Inverse Semigroups)

A semigroup is inverse if and only if it is regular and its idempotents commute.

In an inverse semigroup S the idempotents form a meet semilattice E_S.

Proof An implication, due to V.V. Vagner [Va2], in 1952, is easy: we assume that the semigroup S is regular and its idempotents commute, and prove the uniqueness of partial inverses.

Indeed, if b and c are partial inverses to a, then the products ab, ba, ac, ca are (commuting) idempotents, and b must coincide with c

$$\begin{aligned} b = bab = b(aca)b = b(a(cacac)a)b = (ba)(ca)c(ac)(ab) \\ = (ca)(ba)c(ab)(ac) = c(aba)c(aba)c = cacac = cac = c. \end{aligned}$$

The other implication, due to A.E. Liber [Li], in 1954, is harder to prove. We suppose that S is an inverse semigroup, take two idempotents e, f and show that they commute, following the argument in Clifford–Preston [CP], Theorem 1.17.

First we prove that the product ef is idempotent. Let a be the partial inverse of ef

$$(ef)a(ef) = ef, \qquad a(ef)a = a.$$

Then $b = ae$ is also a partial inverse of ef, because

$$\begin{aligned} (ef)b(ef) = (ef)ae(ef) = (ef)a(ef) = ef, \\ b(ef)b = ae(ef)ae = a(ef)ae = ae. \end{aligned}$$

It follows that $a = ae$. Similarly $fa = a$, and then a is idempotent:

$$aa = ae.fa = a(ef)a = a.$$

Therefore a is the partial inverse of itself, and coincides with ef, which is idempotent; similarly fe is idempotent. Finally, ef (being idempotent) is the partial inverse of itself; but it is also the partial inverse of fe

$$(ef)(fe)(ef) = ef.ef = ef, \qquad (fe)(ef)(fe) = fe.fe = fe,$$

and therefore $ef = fe$.

The last claim follows from (1.11): in any semigroup the product of two commuting idempotents is idempotent. □

1.2.7 Orthodox semigroups

A semigroup S is said to be *orthodox* if it is regular and its idempotents are closed under product: $E_S.E_S \subset E_S$. The name was introduced by Hall [Ha] in 1969, but the property had already been used under different names, by Schein [Sc] and Reilly–Scheiblich [ReS].

Every inverse semigroup is orthodox, by the previous theorem. On the other hand, every orthodox semigroup S has a finest 'inverse congruence' (making the quotient an inverse semigroup), as proved in [Ha].

A well-known characterisation, that will not be used here, says that a regular semigroup is orthodox if and only if every partial inverse of an idempotent element is also idempotent ([ReS], Lemma 1.3).

Orthodox categories, a natural extension of orthodox semigroups, are a main tool in the author's analysis of coherence in Homological Algebra [G9].

1.2.8 Lemma (Regularity Lemma for semigroups)

In the semigroup S we assume that:

$$a = aba, \qquad a' = a'b'a'. \tag{1.19}$$

Then the following properties are equivalent:

(i) $a'a = (a'a)(bb')(a'a)$,
(ii) the element $(b'a')(ab)$ is idempotent.

Note. A similar result can be found in [ReS], Lemma 1.1. Note also that the element $(b'a')(ab)$ is a product of idempotents; therefore *condition* (ii) *is automatically satisfied when S is an orthodox semigroup*, or more particularly an inverse semigroup.

Proof The fact that (i) implies (ii) is trivial (and independent of (1.19)):

$$(b'a'ab)(b'a'ab) = b'(a'a)(bb')(a'a)b = b'(a'a)b.$$

Conversely, if (ii) holds:

$$\begin{aligned}(a'a)(bb')(a'a) &= a'(abb'a')a = (a'b'a')(abb'a')(aba) \\ &= a'(b'a'ab)(b'a'ab)a = a'(b'a'ab)a = (a'b'a')(aba) = a'a.\end{aligned}$$

□

1.3 The involution and order of an inverse semigroup

We introduce involutive semigroups (in 1.3.1) and ordered semigroups (in 1.3.7).

An inverse semigroup has a canonical involution, given by the partial inverses, and a canonical order, generated by the order of its idempotent elements (see 1.1.7).

1.3.1 Involutive semigroups

An *involutive semigroup* will be a semigroup S equipped with an endomapping $a \mapsto a^\sharp$ that reverses the product and is involutive

$$(ab)^\sharp = b^\sharp.a^\sharp, \qquad (a^\sharp)^\sharp = a. \tag{1.20}$$

This mapping is an anti-automorphism of S, that is an isomorphism $S \to S^{\mathrm{op}}$. It automatically preserves the unit 1, if it exists: in fact

$$1^\sharp.a = 1^\sharp.a^{\sharp\sharp} = (a^\sharp 1)^\sharp = a^{\sharp\sharp} = a,$$

for every $a \in S$.

The involution will be said to be *regular* (in the sense of von Neumann), if for every $a \in S$

$$aa^\sharp a = a. \tag{1.21}$$

Then we also have $a^\sharp a a^\sharp = a^\sharp$, so that a and $a^\sharp$ are partial inverses and the semigroup S is regular (cf. 1.2.1). Here we are only interested in regular involutions, that play a central role in categories of partial bijections or categories of relations of modules (and their extensions).

1.3.2 Projectors

Let S be a semigroup with regular involution. A *projector* of S will be a symmetric idempotent, i.e. an element e such that $ee = e = e^\sharp$.

They form a set

$$\mathsf{Prj}\,(S) \subset E_S. \tag{1.22}$$

In [G3]–[G5] these elements are called 'projections', a name with too many meanings.

Every element $a \in S$ has two associated projectors, which will be written as follows:

$$\begin{aligned} &\underline{e}(a) = a^\sharp a && \text{(the } \textit{support} \text{ of } a), \\ &\underline{e}^*(a) = aa^\sharp = \underline{e}(a^\sharp) && \text{(the } \textit{cosupport} \text{ of } a). \end{aligned} \tag{1.23}$$

This terminology is explained by its extension to categories, where the support (resp. cosupport) of the morphism a is a projector of its domain (resp. codomain), cf. 2.2.2.

We have

$$a = a.\underline{e}(a) = \underline{e}^*(a).a = \underline{e}^*(a).a.\underline{e}(a). \tag{1.24}$$

All the projectors of S arise as supports (and cosupports): an element $e \in S$ is a projector if and only if $e = \underline{e}(e)$, if and only if $e = \underline{e}^*(e)$.

The product of two projectors e, f is always idempotent

$$ef.ef = ef.f^\sharp e^\sharp.ef = ef.(ef)^\sharp.ef = ef, \tag{1.25}$$

and is a projector if and only if e and f commute (because $(ef)^\sharp = f^\sharp e^\sharp = fe$).

Conversely every idempotent e is the product of two projectors:

$$e = ee^\sharp e = ee^\sharp.e^\sharp e = \underline{e}^*(e).\underline{e}(e). \tag{1.26}$$

1.3.3 Endo-relations of modules

If A is an R-module, the endo-relations a of A (i.e. the submodules of $A \oplus A$) form a unital semigroup $\mathcal{R}(A)$. Their product is the composition of relations, defined by the following formula

$$ba = \{(x, z) \in A \oplus A \mid \exists\, y \in A\colon (x, y) \in a \text{ and } (y, z) \in b\}. \tag{1.27}$$

This monoid has an obvious involution, given by reversing pairs:

$$a^\sharp = \{(x, y) \in A \oplus A \mid (y, x) \in a\}, \tag{1.28}$$

and *this involution is easily proved to be regular* (see Exercise (a), below). $\mathcal{R}(A)$ is thus a regular monoid.

Exercises and complements. (a) The involution (1.28) is regular.

(b) The endo-relations of a set X, i.e. the subsets of $X \times X$, form a semigroup with a similar product and involution; the latter is not regular, provided that X has at least two elements.

*(c) As a consequence of a Coherence Theorem in Homological Algebra ([G9], Theorem 2.7.6), the following conditions on an R-module A are equivalent:

- the (regular) monoid $\mathcal{R}(A)$ of endo-relations of A is orthodox,

- the (modular) lattice $\mathrm{Sub}(A)$ of submodules of A is distributive,

- the canonical isomorphisms between the subquotients of A are closed under composition, and form a coherent system of isomorphisms.

1.3.4 Proposition and Definition

Let S be an inverse semigroup.

(a) S has a canonical involution*: the mapping $a \mapsto a^\sharp$ that takes each element to its partial inverse. This is the unique regular involution of S.*

(b) The idempotents of S are the same as its projectors for the involution. They commute and form a meet semilattice $E_S = \mathsf{Prj}\,(S)$

$$e \leqslant f \;\Leftrightarrow\; e = ef = fe, \qquad e \wedge f = ef = fe. \tag{1.29}$$

Proof (a) Plainly, if S has a regular involution $a \mapsto a^\sharp$, then $a^\sharp$ must be the partial inverse of a.

Conversely, defining $a^\sharp$ in this way, we have to verify that the partial inverse of ab is $b^\sharp a^\sharp$; this follows easily from the fact that the idempotents of S commute (by Theorem 1.2.6):

$$(ab)(b^\sharp a^\sharp)(ab) = a(bb^\sharp)(a^\sharp a)b = a(a^\sharp a)(bb^\sharp)b = ab.$$

Similarly $(b^\sharp a^\sharp)(ab)(b^\sharp a^\sharp) = b^\sharp a^\sharp$.

(One can also deduce these facts from the Regularity Lemma 1.2.8.)

(b) In an inverse semigroup every idempotent e is its own partial inverse (as already remarked in 1.2.3), and therefore a symmetric element with respect to the canonical involution: $e^\sharp = e$.

The rest follows from the commutativity of idempotents. □

1.3.5 Proposition (Characterisation of inverse semigroups, II)

For a semigroup S the following conditions are equivalent:

(i) S is inverse,

(ii) S has a regular involution and its idempotents commute,

(iii) S has a regular involution and its projectors commute,

(iv) S has a regular involution and its projectors form a subsemigroup.

Proof It is an easy consequence of Theorem 1.2.6. Indeed, the equivalence of (i) and (ii) follows from 1.2.6 and 1.3.4 (that ensures the existence of a regular involution in every inverse semigroup).

Then, condition (ii) trivially implies (iii). The converse follows from the fact that, once we have a regular involution, every idempotent is the product of two projectors (as we have seen in 1.3.2). The equivalence of (iii) and (iv) is obvious, since $(ef)^\sharp = f^\sharp e^\sharp$. □

1.3.6 *Theorem (Vagner-Preston Representation Theorem)

Every inverse semigroup S can be embedded as an inverse subsemigroup of the symmetric inverse semigroup $\mathcal{I}(S)$ (over its underlying set, see 1.2.4), by the left translation:

$$\begin{aligned} \lambda\colon S \to \mathcal{I}(S), \qquad & a \mapsto \lambda_a, \\ \lambda_a\colon a^\sharp S \to aS, \qquad & \lambda_a(x) = ax \quad (\text{for } a \in S). \end{aligned} \tag{1.30}$$

Note. In other words, every inverse semigroup is a semigroup of partial bijections, as proved by Vagner [Va2] in 1952 and independently by Preston in 1954. This result extends Cayley's theorem: every group is a group of bijections. It will not be used here.

Proof We follow the argument in Clifford–Preston [CP], Theorem 1.20, with a slightly different notation.

First, let us note that the subset $a^\sharp S = a^\sharp aS$ admits the idempotent $\underline{e}(a) = a^\sharp a$ as a left unit, while the subset $aS = aa^\sharp S$ admits the idempotent $\underline{e}^*(a) = aa^\sharp$ as a left unit.

This shows that $\lambda_a\colon a^\sharp S \to aS$ is a bijective mapping, with inverse $\lambda_{a^\sharp}\colon aS \to a^\sharp S$.

Globally, the mapping $\lambda\colon S \to \mathcal{I}(S)$ is injective. In fact, if $\lambda_a = \lambda_b$ we have $a^\sharp aS = b^\sharp bS$; but then $a^\sharp a = (a^\sharp a)(a^\sharp a) = (b^\sharp b)y$ (for some $y \in S$) and $(b^\sharp b)(a^\sharp a) = a^\sharp a$, which means that $a^\sharp a \leqslant b^\sharp b$ in the meet semilattice $E = \mathsf{Prj}\,(S)$; symmetrically, $b^\sharp b \leqslant a^\sharp a$ and these projectors coincide. Now

$$a = aa^\sharp a = \lambda_a(a^\sharp)a = \lambda_b(a^\sharp)a = ba^\sharp a = bb^\sharp b = b.$$

Finally, we have to prove that $\lambda_{ba} = \lambda_b\lambda_a$, for all a, b in S. Since $bax = b(ax)$, it suffices to show that $\mathrm{Val}\,(\lambda_{ba}) = \mathrm{Val}\,(\lambda_b\lambda_a)$ (and recall that $\mathrm{Def}\,(\lambda_a) = \mathrm{Val}\,(\lambda_{a^\sharp})$).

If $z \in \mathrm{Val}\,(\lambda_b\lambda_a)$, there is some $x \in a^\sharp S$ such that $ax \in b^\sharp S$ and $z = bax$; but then $z \in baS = \mathrm{Val}\,(\lambda_{ba})$. Conversely, let $z \in \mathrm{Val}\,(\lambda_{ba}) = baS = baa^\sharp b^\sharp S$. Then $z = ba(a^\sharp b^\sharp z')$ and the element $x = a^\sharp b^\sharp z'$ belongs to $a^\sharp S$. Its image

$$\lambda_a(x) = aa^\sharp b^\sharp z' = (aa^\sharp)(b^\sharp b)b^\sharp z' = (b^\sharp b)(aa^\sharp)b^\sharp z',$$

belongs to $b^\sharp S$; thus $\lambda_b(\lambda_a(x)) = ba(a^\sharp b^\sharp z') = z$ belongs to $\mathrm{Val}\,(\lambda_b\lambda_a)$. □

1.3.7 Ordered semigroups

An *ordered semigroup* is a semigroup S equipped with an order relation $a \leqslant b$ consistent with the product:

$$\text{if } a \leqslant a' \text{ and } b \leqslant b' \text{ in } S, \text{ then } ab \leqslant a'b'. \tag{1.31}$$

Equivalently, for every a and $b \leqslant b'$ in S, we have: $ab \leqslant ab'$ and $ba \leqslant b'a$.

Examples. In a meet (or join) semilattice the order is always consistent with the operation. In an ordered ring the sum is consistent with the ordering, but only the elements $a \geqslant 0$ give a monotone multiplication:

$$\text{if } b \leqslant b' \text{ in } S, \text{ then } ab \leqslant ab' \text{ and } ba \leqslant b'a.$$

1.3.8 Theorem

Let S be an inverse semigroup and $E = \mathsf{Prj}\,(S)$ its meet semilattice of idempotents.

For $a, b \in S$ the following properties are equivalent:

(*i*) $a = ab^\sharp a$,
(*ii*) $a = b(a^\sharp a)$,
(*ii**) $a = (aa^\sharp)b$,
(*iii*) $a = (aa^\sharp)b(a^\sharp a)$,
(*iv*) $a \in bE$,
(*iv**) $a \in Eb$,
(*v*) $a \in EbE$.

Proof Trivially, (ii) implies (iv). Conversely, if $a = be$ for some $e \in E$, then

$$b(a^\sharp a) = b(eb^\sharp be) = be(b^\sharp b)e = bb^\sharp b.e = be = a.$$

Similarly (ii*) is equivalent to (iv*) and (iii) is equivalent to (v).

Now, (iv) implies (v): if $a = be$ (with $e \in E$) then $a = (bb^\sharp)be \in EbE$. Conversely, if $a = fbe$ (with $e, f \in E$) then

$$a = f(bb^\sharp)be = (bb^\sharp)fbe = b(b^\sharp fb)e$$

belongs to bE, because $b^\sharp fb$ is idempotent:

$$(b^\sharp fb)(b^\sharp fb) = b^\sharp f(bb^\sharp)fb = b^\sharp ff(bb^\sharp)b = b^\sharp fb.$$

Similarly, (iv*) is equivalent to (v), and we are left with proving that (i) is equivalent to the other properties.

The fact that (i) implies (v) is easily seen: if $a = ab^\sharp a$, then $a = a(b^\sharp bb^\sharp)a = (ab^\sharp)b(b^\sharp a)$ belongs to EbE.

Conversely, let $a = fbe$ (with e, f idempotents). Applying the Regularity Lemma 1.2.8, from $e = eee$ and $fb = (fb)(b^\sharp f)(fb)$ we deduce that $fb.e = (fbe)(eb^\sharp f)(fbe)$. Finally:

$$ab^\sharp a = (fbe)b^\sharp(fbe) = (fbe)(eb^\sharp f)(fbe) = fbe = a.$$

□

1.3.9 Theorem and Definition

(a) The inverse semigroup S has a canonical order $a \leqslant b$ *(consistent with the product and partial inverses), that is defined by the equivalent properties (i)–(v) of the previous theorem.*

(b) This order of S extends the order in the meet semilattice $E = \mathsf{Prj}\,(S)$ of idempotent elements, defined in 1.1.7, and is generated by the latter (as an order of semigroups).

(c) If S is unital, the idempotent elements e are characterised by the condition $e \leqslant 1$. Furthermore, the semigroup-order of S is generated by the condition $e \leqslant 1$, for all idempotents e.

Proof (a) The relation $a \leqslant b$, is reflexive (use property 1.3.8(i)) and transitive (use 1.3.8(iv)); it is also antisymmetric: if $a = be$ and $b = af$ (with $e, f \in E$), then $a = afe = aef = be.ef = af = b$.

The fact that the order is consistent with the product follows from the Regularity Lemma 1.2.8. Indeed, assuming that $a \leqslant b$ and $a' \leqslant b'$, we can apply the lemma to the pairs $(a, b^\sharp)$ and $(a', b'^\sharp)$, concluding that $aa' = (aa')(b'^\sharp b^\sharp)(aa')$, which means that $aa' \leqslant bb'$.

Finally, if $a \leqslant b$, then $a^\sharp \leqslant b^\sharp$.

(b) The first claim is obvious. As to the second, let us suppose we have in S a semigroup-order $a \prec b$ that extends the order $\leqslant$ of E and prove that the canonical order is finer.

If the pair a, b satisfies the equivalent properties (i)–(v) of 1.3.8, then $ab^\sharp$ and $b^\sharp a$ are idempotent and

$$ab^\sharp = (aa^\sharp)(bb^\sharp) \leqslant bb^\sharp, \qquad b^\sharp a = (b^\sharp b)(a^\sharp a) \leqslant b^\sharp b.$$

Therefore, since the semigroup-order $\prec$ extends the order of projectors:

$$a = ab^\sharp a \prec bb^\sharp a \prec bb^\sharp b = b.$$

(c) An obvious consequence of property 1.3.8(iv). □

1.4 Categories

In the second part of this chapter we review the basic notions of category theory. Examples and exercises will focus on categories of topological spaces or ordered sets, rather than algebraic structures — less present in this analysis of local structures.

The theory of categories was established by Eilenberg and Mac Lane in 1945, in a well-known paper [EiM]. The interested reader will explore with pleasure the books of S. Mac Lane [M3] and F. Borceux [Bo].

Two earlier books by B. Mitchell and P. Freyd [Mi, Fr1] are centred on abelian categories and their embedding in categories of modules. The book by J. Adámek, H. Herrlich and G.E. Strecker [AHS] gives an accurate analysis of 'concrete categories'.

At a more elementary level, the author's [G10] is a textbook for beginners, also devoted to applications in Algebra, Topology and Algebraic Topology.

Dealing with categories, one should avoid the usual paradoxes related to 'the set of all sets'. Here we make use of a particular set theory where there are *sets* and *classes*; every set is also a class, but a *proper* class is not a set: for instance the *class of all sets* and the *class of all groups* are proper (see [AHS] or the Appendix of [Ke]). This approach is followed in [Mi, Fr1, AHS].

When two levels, like sets and classes, are insufficient, one can introduce a third level of 'hyperclasses', called 'conglomerates' in [AHS].

*Alternatively, a more flexible setting widely used for categories is ordinary Set Theory (say ZFC), with the assumption of the existence of a Grothendieck *universe*, or of a suitable hierarchy of universes (cf. [M3], Section I.6).*

1.4.1 Categories of structured sets

Loosely speaking, before giving a precise definition, a category C consists of *objects* and *morphisms*, together with a (partial) *composition law*: given three objects X, Y, Z and two 'consecutive' morphisms $f\colon X \to Y$ and $g\colon Y \to Z$, we have a composed morphism $gf\colon X \to Z$.

This partial operation is associative (whenever composition is legitimate); every object X has an identity $\mathrm{id}\,X$ which acts as a unit for legitimate compositions.

The prime example is the category Set *of sets* (and mappings), where:

- an object is a set,

- the morphisms $f\colon X \to Y$ between two given sets X and Y are the (set-theoretical) mappings from X to Y,

- the composition law is the usual composition of mappings: $(gf)(x) = g(f(x))$.

The following categories 'of structured sets' will often be used:

- the category Top *of topological spaces* (and continuous mappings),

- the category Ord of *ordered sets* (and monotone mappings),

- the category pOrd of *preordered sets* (and monotone mappings),

- the category $\mathsf{Set}_\bullet$ *of pointed sets* (and pointed mappings),

- the category $\mathsf{Top}_\bullet$ *of pointed topological spaces* (and pointed maps),

- the category Ab *of abelian groups* (and homomorphisms),

- the category Gp *of groups* (and homomorphisms),

- the category $R\,\mathsf{Mod}$ *of (unitary) left modules* on a given unital ring R,

- the category Ban of *Banach spaces and continuous linear mappings*,

- the category Ban_1 of *Banach spaces and linear weak contractions.*

For the category $\mathsf{Set}_\bullet$, let us recall that a *pointed set* is a pair (X, x_0) consisting of a set X and a *base*-element $x_0 \in X$, while a pointed mapping $f\colon (X, x_0) \to (Y, y_0)$ is a mapping $f\colon X \to Y$ such that $f(x_0) = y_0$; their composition is obvious.

Similarly, a *pointed topological space* (X, x_0) is a space with a base-point, and a pointed map $f\colon (X, x_0) \to (Y, y_0)$ is a continuous mapping from X to Y such that $f(x_0) = y_0$. The reader may know that the category $\mathsf{Top}_\bullet$ is important in Algebraic Topology: for instance, the fundamental group $\pi_1(X, x_0)$ is defined for a pointed topological space.

The categories Ban and Ban_1 will be used in a marginal, elementary way, for examples and counterexamples. It is understood that we have chosen *either* the real *or* the complex field of scalars. We also recall that a linear weak contraction is a linear mapping with norm $\leqslant 1$.

A reader interested in the categorical aspects of Banach spaces is referred to Semadeni's book [Se].

When a category is named after its objects alone (e.g. the 'category of groups'), this means that the morphisms are understood to be the obvious ones (in this case the group-homomorphisms) and the composition is understood to be the usual one.

Of course, different categories with the same objects have different names, like Ban and Ban_1.

1.4.2 Definition

A *category* C consists of the following data:

(a) a class $\mathrm{Ob}\,\mathsf{C}$, whose elements are called *objects* of C,

(b) for every pair X, Y of objects, a set $\mathsf{C}(X, Y)$ (called a *hom-set*) whose elements are called *morphisms* (or *maps*, or *arrows*) of C from X to Y and written as $f\colon X \to Y$,

(c) for every triple X, Y, Z of objects of C, a mapping of sets, called *composition*

$$\mathsf{C}(X, Y) \times \mathsf{C}(Y, Z) \to \mathsf{C}(X, Z), \qquad (f, g) \mapsto gf,$$

that gives a *partial composition law* for pairs of consecutive morphisms. The composite gf will also be written as $g.f$.

These data must satisfy the following axioms.

(i) (*Associativity*) Given three consecutive arrows, $f\colon X \to Y$, $g\colon Y \to Z$ and $h\colon Z \to W$, one has: $h(gf) = (hg)f$.

(ii) (*Identities*) Given an object X, there exists an endo-map $e\colon X \to X$ which acts as a unit whenever composition makes sense; in other words if $f\colon X' \to X$ and $g\colon X \to X''$, one has: $ef = f$ and $ge = g$.

One shows, in the usual way, that e is determined by X; it is called the *identity* of X and written as 1_X or $\mathrm{id}\,X$.

We generally assume that the following condition is also satisfied.

(iii) (*Separation*) Every map $f\colon X \to Y$ has a well-determined *domain* $\mathrm{Dom}\,f = X$ and a well-determined *codomain* $\mathrm{Cod}\,f = Y$.

Concretely, when constructing a category, one can forget about this condition, since one can always satisfy it *redefining* a morphism $\hat{f}\colon X \to Y$ as a triple $(X, Y; f)$ where f is a morphism from X to Y in the original sense (possibly not satisfying the Separation axiom).

Two morphisms f, g are said to be *parallel* when they have the same domain and the same codomain; the meaning of a pair of *consecutive* arrows has already been mentioned.

$\mathrm{Mor}\,\mathsf{C}$ denotes the class of all the morphisms of C, i.e. the disjoint union of its hom-sets. The endomorphisms of every object X form a monoid

$$\mathsf{C}(X) = \mathsf{C}(X, X). \tag{1.32}$$

If C is a category, the *opposite* (or *dual*) category, written as C^{op}, has the same objects as C and 'reversed arrows', with 'reversed composition' $g * f$

$$\mathsf{C}^{\mathrm{op}}(X, Y) = \mathsf{C}(Y, X), \qquad g * f = fg, \qquad \mathrm{id}^{\,\mathrm{op}}(X) = \mathrm{id}\,X. \tag{1.33}$$

Every topic of category theory has a dual instance, which comes from the opposite category (or categories). A dual notion is often distinguished by the prefix 'co-'.

1.4.3 Small and large categories

We have assumed that a category C has a *class* $\mathrm{Ob}\,\mathsf{C}$ of objects (e.g. the class of all sets, or the class of all topological spaces) and, for every pair X, Y of objects, a *set* $\mathsf{C}(X, Y)$ of morphisms from X to Y.

The categories of structured sets that we consider, like the examples of 1.4.1, are generally *large* categories, where the objects (or equivalently the morphisms) form a *proper class* (i.e. not a set). A category C is said to be *small* if the class $\mathrm{Ob}\,\mathsf{C}$ is a set. A *finite* category is a small category whose set of morphisms is finite (then the same is true of its set of objects, since an object is determined by its identity).

A set X can be viewed as a *discrete* small category: its objects are the elements of X, and the only arrows are their (formal) identities.

A preordered set X will often be viewed as a small category, where the objects are the elements of X and the set $X(x, x')$ contains precisely one arrow if $x \prec x'$ (which can be written as $(x, x')\colon x \to x'$), and no arrow otherwise. Composition and identities are uniquely determined, as follows

$$(x', x'').(x, x') = (x, x''), \qquad \mathrm{id}\, x = (x, x), \tag{1.34}$$

and all diagrams in such a category commute.

In this sense, categories generalise preordered sets. Their dualities agree: the category associated to the opposite preordered set X^{op} (in 1.1.1) is dual to the category associated to X. Loosely speaking, the extension from preordered sets (or classes) to categories consists in allowing different arrows between specified objects.

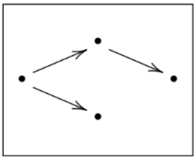
In particular, each finite ordinal defines a category, which is often written as $\mathbf{0}, \mathbf{1}, \mathbf{2}, ...$ Thus, $\mathbf{0}$ is the empty category; $\mathbf{1}$ is the *singleton category*, i.e. the discrete category on one object; $\mathbf{2}$ is the *arrow category*, with two objects (0 and 1), and one non-identity arrow, $0 \to 1$.

A monoid M can be viewed as a small category M with one object, say $*$, and set of morphisms $\mathrm{Mor}\,\mathsf{M} = \mathsf{M}(*) = M$, with composition the multiplication of M. *In this sense, categories generalise monoids.* Again, their dualities coincide (see 1.1.5). Loosely speaking, a category can be viewed as a 'many-object' extension of a monoid.

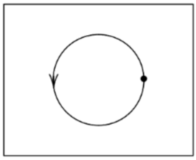
The relationship between categories and semigroups will be further explored in 2.3.1 and Section 2.8.

1.4.4 Structural remarks

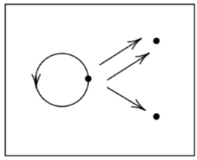
The following diagrams represent a preordered set and a monoid as small categories of different kinds (as remarked above), and how a category can combine both structures



a preordered set



a monoid



a category

We only draw generators: composed arrows and identities are 'virtually present'. Thus, the second diagram represents a monoid with one generator.

We shall see how some aspects of the theory of preordered sets and the theory of semigroups can be extended to categories:

(a) (*many-arrow extensions*) infima and suprema extend to *categorical products* and *sums*, Galois connections to *adjunctions*,

(b) (*many-object extensions*) regular and inverse monoids extend to *von Neumann regular* and *inverse* categories.

In a very loose sense, the alternative between (a) and (b) is concerned with aspects of category theory where the objects play a relevant role, or not.

Of course, each of the three theories has its specific topics, whose extension or restriction can be of little interest.

1.4.5 Isomorphisms

In a category C a morphism $f\colon X \to Y$ is said to be *invertible*, or an *isomorphism*, if it has an inverse, i.e. a morphism $g\colon Y \to X$ such that $gf = 1_X$ and $fg = 1_Y$. The latter is uniquely determined by f and written as f^{-1}.

The *isomorphism relation* $X \cong Y$ between objects of C (meaning that there exists an isomorphism $X \to Y$) is obviously an equivalence relation.

For instance, the isomorphisms of Set, Top and Ab are, respectively: the bijective mappings of sets, the homeomorphisms of topological spaces and the isomorphisms of abelian groups. The isomorphisms of Ban are the linear homeomorphisms; the isomorphisms of Ban_1 are more restricted, namely the bijective linear isometries.

A *groupoid* is a category where every map is invertible. As above, a groupoid can be viewed as 'a group on many objects'.

Every category C has an *associated groupoid* $\mathrm{Iso}(\mathsf{C})$ on the same objects: its arrows are the isomorphisms of C, composed as in C.

Examples and complements. (a) If X is a preordered set, viewed as a category, two elements x, x' are isomorphic objects if and only if they correspond each other in the associated equivalence relation (in 1.1.1). The groupoid $\mathrm{Iso}(X)$ is this equivalence relation, viewed as a category.

(b) If M is a monoid, viewed as a category, an isomorphism amounts to an invertible element, and $\mathrm{Iso}(M)$ is their group.

(c) Groupoids were introduced before categories, by H. Brandt in 1927 [Bra]. Brandt dealt with 'connected' groupoids, in a single-sorted version where the only terms are the arrows, and the objects are represented by their identities. Categories can also be presented in a single-sorted version: see [M3], p. 9.

*(d) In the *fundamental groupoid* $\Pi_1(X)$ of a topological space X, an object is any point $x \in X$ and an arrow $[a]\colon x \to y$ is a class of paths in X, from x to y, up to homotopy with fixed end-points. The construction is sketched in 4.3.1(b).

1.4.6 Subcategories, quotients and products of categories

(a) Let C be a category. A *subcategory* D is defined by assigning:

- a subclass $\mathrm{Ob}\,\mathsf{D} \subset \mathrm{Ob}\,\mathsf{C}$, whose elements are called *objects of* D,

- for every pair of objects X, Y of D, a subset $\mathsf{D}(X, Y) \subset \mathsf{C}(X, Y)$, whose elements are called *morphisms of* D, from X to Y,

so that the following conditions hold:

(i) the composite in C of morphisms of D belongs to D,

(ii) the identity in C of an object of D belongs to D.

Then D, equipped with the restricted composition law, is a category.

We say that D is a *full* subcategory of C if, for every pair of objects X, Y of D, we have $\mathsf{D}(X, Y) = \mathsf{C}(X, Y)$, so that D is determined by assigning a subclass of objects. We say that D is a *wide* subcategory of C if it has the same objects.

For instance, Ab is a full subcategory of Gp, Ord is a full subcategory of pOrd, while Ban_1 is a wide subcategory of Ban and $\mathrm{Iso}(\mathsf{C})$ of C. Of course a full and wide subcategory must be the total one.

(b) A *congruence* $R = (R_{XY})$ in a category C consists of a family of equivalence relations R_{XY} in each set of morphisms $\mathsf{C}(X, Y)$, that is consistent with composition:

$$\text{if } f\, R_{XY}\, f' \text{ and } g\, R_{YZ}\, g', \text{ then } (gf)\, R_{XZ}\, (g'f'). \tag{1.35}$$

Then one defines the *quotient category* $\mathsf{D} = \mathsf{C}/R$: the objects are those of C, and $\mathsf{D}(X, Y) = \mathsf{C}(X, Y)/R_{XY}$; in other words, a morphism $[f]\colon X \to Y$ in D is an equivalence class of morphisms $X \to Y$ in C. The composition is induced by that of C, which is legitimate because of condition (1.35):

$$[g].[f] = [gf]. \tag{1.36}$$

For instance, in Top the homotopy relation $f \simeq f'$ is (well-known to be) a congruence of categories; the quotient category $\mathsf{hoTop} = \mathsf{Top}/\simeq$ is called the *homotopy category of topological spaces*, and is important in Algebraic Topology. Plainly, a continuous mapping $f\colon X \to Y$ is a homotopy equivalence if and only if its homotopy class $[f]$ is an isomorphism of the category hoTop.

The relation of isomorphism is wider in a quotient category (but it may coincide with the original one, also in a non-trivial quotient).

(c) If C and D are categories, one defines the *product category* $\mathsf{C} \times \mathsf{D}$. An object is a pair (X, Y) where X is in C and Y in D; a morphism is a pair of morphisms

$$(f, g)\colon (X, Y) \to (X', Y'), \qquad (f \in \mathsf{C}(X, X'),\ g \in \mathsf{D}(Y, Y')). \tag{1.37}$$

The composition of (f, g) with a consecutive morphism

$$(f', g')\colon (X', Y') \to (X'', Y'')$$

is (obviously) defined component-wise: $(f', g').(f, g) = (f'f, g'g)$.

More generally one defines the *cartesian product* $\mathsf{C} = \prod_{i \in I} \mathsf{C}_i$ of a family of categories $(\mathsf{C}_i)_{i \in I}$ indexed by a set I: an object of C is a family $(A_i)_{i \in I}$ where $A_i \in \mathrm{Ob}\,(\mathsf{C}_i)$ (for every index i), and a morphism $f = (f_i)\colon (A_i) \to (B_i)$ is a family of morphisms $f_i \in \mathsf{C}_i(A_i, B_i)$; the composition is component-wise and $\mathrm{id}\,((A_i)_{i \in I}) = (\mathrm{id}\, A_i)_{i \in I}$.

1.4.7 Monomorphisms and epimorphisms

In a category, monomorphisms and epimorphisms, are defined by cancellation properties with respect to composition. In a category of structured sets, they represent an 'approximation' to the injective and surjective mappings of the category, and may coincide or not with the latter.