Alan J. Cain

Nine Chapters on the Semigroup Art

Lecture notes for a tour through semigroups

Porto & Lisbon
2020



version 0.66.62 (2020-06-13) [A4 / Two-sided / Colour]

3b9596955641bc1af1e88645064dd8dc6c2cf041 : 000000005ee533ff

To download the most recent version, and files suitable for colour or greyscale printing, or for viewing on tablets and ebook readers, visit

http://www-groups.mcs.st-andrews.ac.uk/~alanc/pub/c_semigroups/

© 2012–20 Alan J. Cain (a.cain@fct.unl.pt)



This work is licensed under the Creative Commons Attribution–Non-Commercial–NoDerivs 4.0 International Licence. To view a copy of this licence, visit

https://creativecommons.org/licenses/by-nc-nd/4.0/

Contents

Preface v

Prerequisites vii ◆ Acknowledgements vii

Chapter 1 | Elementary semigroup theory 1

Basic concepts and examples 1 ◆ Generators and subsemigroups 8 ◆ Binary relations 11 ◆ Orders and lattices 15 ◆ Homomorphisms 19 ◆ Congruences and quotients 20 ◆ Generating equivalences and congruences 22 ◆ Subdirect products 28 ◆ Actions 29 ◆ Cayley graphs 30 ◆ Exercises 32 ◆ Notes 34

Chapter 2 | Free semigroups & presentations 37

Alphabets and words 37 ◆ Universal property 38 ◆ Properties of free semigroups 41 ◆ Semigroup presentations 42 ◆ Exercises 51 ◆ Notes 53

Chapter 3 | Structure of semigroups 55

Green's relations 55 ◆ Simple and 0-simple semigroups 58 ◆ $\mathcal{D}$-class structure 60 ◆ Inverses and $\mathcal{D}$-classes 63 ◆ Schützenberger groups 65 ◆ Exercises 68 ◆ Notes 70

Chapter 4 | Regular semigroups 73

Completely 0-simple semigroups 75 ◆ Ideals and completely 0-simple semigroups 81 ◆ Completely simple semigroups 82 ◆ Completely regular semigroups 84 ◆ Left and right groups 86 ◆ Homomorphisms 88 ◆ Exercises 89 ◆ Notes 91

Chapter 5 | Inverse semigroups 93

Equivalent characterizations 93 ◆ Vagner–Preston theorem 97 ◆ The natural partial order 100 ◆ Clifford semigroups 102 ◆ Free inverse semigroups 106 ◆ Exercises 116 ◆ Notes 119

Chapter 6 | Commutative semigroups 121

Cancellative commutative semigroups 121 ◆ Free commutative semigroups 123 ◆ Rédei's theorem 125 ◆ Exercises 128 ◆ Notes 129

Chapter 7 | Finite semigroups 131
Green's relations and ideals 131 ◆ Semidirect and wreath products 133 ◆ Division 135 ◆ Krohn–Rhodes decomposition theorem 140 ◆ Exercises 147 ◆ Notes 148

Chapter 8 | Varieties & pseudovarieties 149
Varieties 149 ◆ Pseudovarieties 157 ◆ Pseudovarieties of semigroups and monoids 159 ◆ Free objects for pseudovarieties 161 ◆ Projective limits 162 ◆ Pro-V semigroups 164 ◆ Pseudoidentities 167 ◆ Semidirect products of pseudovarieties 172 ◆ Exercises 173 ◆ Notes 174

Chapter 9 | Automata & finite semigroups 175
Finite automata and rational languages 175 ◆ Syntactic semigroups and monoids 184 ◆ Eilenberg correspondence 188 ◆ Schützenberger's theorem 195 ◆ Exercises 201 ◆ Notes 202

Solutions to exercises 203

Bibliography 251

Index 255

LIST OF TABLES

Table 8.1 Varieties of semigroups 156
Table 8.2 Varieties of monoids 156
Table 8.3 Varieties of semigroups with a unary operation $^{-1}$ 157
Table 8.4 $\mathcal{S}$-pseudovarieties of semigroups 169
Table 8.5 $\mathcal{M}$-pseudovarieties of monoids 169

Table 9.3 Varieties of rational $*$-languages 193
Table 9.4 Varieties of rational $+$-languages 193

Preface

‘A preface is frequently a superior composition to the work itself’

— Isaac D’Israeli,
‘Prefaces’. In: *Curiosities of Literature*.

This course is a tour through selected areas of semigroup theory. There are essentially three parts:

- Chapters 1–3 study general semigroups, including presentations for semigroups and basic structure theory.
- Chapters 4–6 examine special classes: namely regular, inverse, and commutative semigroups.
- Chapters 7–9 study finite semigroups, their classification using pseudovarieties, and connections with the theory of automata and regular languages.

The course is broad rather than deep. It is *not* intended to be comprehensive: it does not try to study (for instance) structure theory as deeply as Howie, *Fundamentals of Semigroup Theory*, pseudovarieties as deeply as Almeida, *Finite Semigroups and Universal Algebra*, or languages as deeply as Pin, *Varieties of Formal Languages*; rather, it samples highlights from each area. It should be emphasized that there is very little that is original in this course. It is heavily based on the treatments in these and other standard textbooks, as the bibliographic notes in each chapter make clear. The main novelty is in the selection and arrangement of material, the slightly slower pace, and the general policy of avoiding leaving proofs to the reader when the corresponding results are required for later proofs.

Figure P.1 shows the dependencies between the chapters. At the end of each chapter, there are a number of exercises, intended to reinforce concepts introduced in the chapter, and also to explore some related topics that are not covered in the main text. The most important exercises are marked with a star ✶ . Solutions are supplied for all exercises. At the end of each chapter are bibliographic notes, which give sources and suggestions for further reading.


Warnings against potential misunderstandings are marked (like this) with a ‘dangerous bend’ symbol, as per Bourbaki or Knuth.


Important observations that are not potential misunderstandings *per se* are marked with an ‘exclamation’ symbol (like this).

This course was originally delivered to master’s students at the Uni-

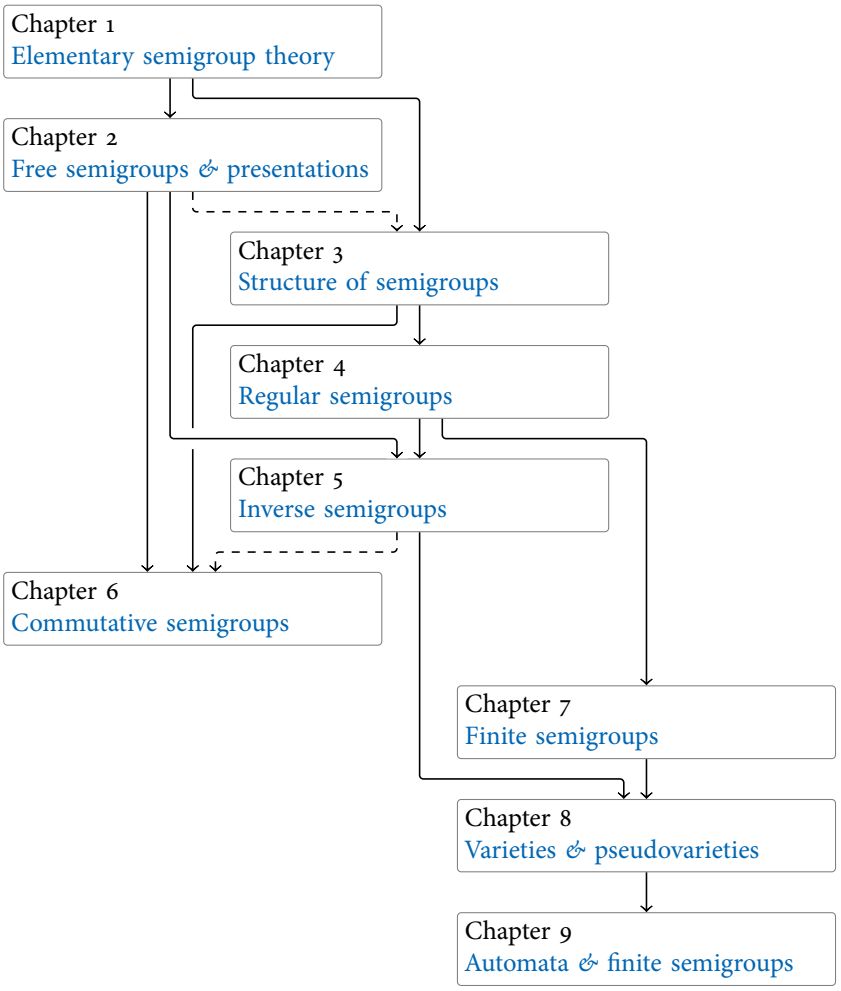


FIGURE P.1
Chart of the dependencies between the chapters. Dotted arrows indicate that the dependency is only in the exercises, not in the main text.

versities of Porto and Santiago de Compostella. The course was covered during 56 hours of classes, which included lectures and discussions of the exercises. Revisions have increased the length of the notes, and about 70 hours of class time would now be required to cover them fully.

These notes were heavily revised in 2013–15. Most of the main text is now stable, but Chapter 8 will be further revised, and further exercises will be added. At present, the index is limited to names and 'named results' only. There may be minor typesetting problems that arise from the 'in-development' status of the LuaLaTeX software and many of the required packages.

The author welcomes any corrections, observations, or constructive criticisms; please send them to the email address on the copyright page.

PREREQUISITES

There are few formal prerequisites: general mathematical maturity is the main one. An understanding of the most basic concepts from elementary group theory is assumed, such as the definition of groups, cosets, and factor groups. Some knowledge of linear algebra will help with understanding certain examples, but is not vital. For Chapters 1 and 5, knowledge of the basic definitions of graph theory is assumed. Some basic topology is necessary to appreciate part of Chapter 8 fully (although most of the chapter can be understood without it, and the relevant sections can simply be skipped), and some background in universal algebra is useful, but not essential. For Chapter 9, some experience with formal language theory and automata is useful, but again not essential.

ACKNOWLEDGEMENTS

Attila Egri-Nagy, Darij Grinberg, Akihiko Koga, and Guilherme Rito made valuable suggestions and indicated various errors. Some exercises were suggested by Victor Maltcev. Typos were pointed out by Nick Ham, Samuel Herman, José Manuel dos Santos dos Santos, and Alexandre Trocado. Many improvements are due to the students who took the first version of this course: Miguel Couto, Xabier García, and Jorge Soares. The imperfections that remain are my responsibility.

The title alludes to 九章算術 (*Jiǔzhāng Suànshù*), *Nine Chapters on the Mathematical Art*.

A. J. C.

Elementary semigroup theory

1

> ‘I use the word “elementary” in the sense in which professional mathematicians use it’
>
> — G.H. Hardy, *A Mathematician’s Apology*, § 21.

A *binary operation* $\circ$ on a set S is a map $\circ : S \times S \to S$. This operation is *associative* if $x \circ (y \circ z) = (x \circ y) \circ z$ for all elements $x, y, z \in S$. A *semigroup* is a non-empty set equipped with an associative binary operation.

Binary operation

Semigroup

Semigroups are therefore one of the most basic types of algebraic structure. We could weaken the definition further by removing the associativity condition and requiring only a binary operation on a set. A structure that satisfies this weaker condition is called a *magma* or *groupoid*. (These ‘groupoids’ are different from the category-theoretic notion of groupoid.)

On the other hand, we can strengthen the definition by requiring an identity and inverses. Structures satisfying this stronger condition are of course *groups*. However, there are many more semigroups than groups. For instance, there are 5 essentially different groups with 8 elements (the cyclic group C_8, the direct products $C_4 \times C_2$ and $C_2 \times C_2 \times C_2$, the dihedral group D_4, and the quaternion group Q_8), but there are 3 684 030 417 different (non-isomorphic) semigroups with 8 elements.



Some authors define a semigroup as a (possibly empty) set equipped with an associative binary operation. That is, the empty set forms the ‘empty semigroup’. This has advantages from a category-theoretic viewpoint. Note, however, that other definitions must be adjusted if a semigroup can be empty. In these notes, semigroups are required to be non-empty.

‘Empty semigroup’

BASIC CONCEPTS AND EXAMPLES

Throughout this chapter, S will denote a semigroup with operation $\circ$. Formally, we write $(S, \circ)$ to indicate that we are considering the set S with the operation $\circ$, but we will only do this when we need to distinguish a particular operation. Unless we need to distinguish between

different operations, we will often write xy instead of $x \circ y$ (where $x, y \in S$) and we will call the operation *multiplication* and the element xy (i.e. the result of applying the operation to x and y) the *product* of the elements x and y.

Multiplication, product

In order to compute a product like $xyzt$ (or, equivalently, $x \circ y \circ z \circ t$), where $x, y, z, t \in S$, we have to insert balanced pairs of brackets into the product to show in what order we perform the multiplications. We might insert brackets in any of the following five ways:

$$((xy)z)t, \quad (x(yz))t, \quad (xy)(zt), \quad x((yz)t), \quad x(y(zt)).$$

The following result shows that the choice of how to insert balanced pairs of brackets is unimportant:

Associativity

PROPOSITION 1.1. *Let $s_1, \ldots, s_n \in S$. Every way of inserting balanced pairs of brackets into the product $s_1 s_2 \cdots s_n$ gives the same result.*

Proof of 1.1. We will prove that any insertion of brackets into the product gives the same result as $s_1(s_2(s_3 \cdots s_n)\cdots)$. We proceed by induction on n. For $n = 1$, the result is trivially true, for there is only one way to insert balanced pairs of brackets into the product s_1. This is the base case of the induction.

So assume that the result holds for all $n < k$; we aim to show it is true for $n = k$. Take some bracketing of the product $s_1 s_2 \cdots s_k$ and let t be the result. This bracketing is a product of some bracketing of $s_1 \cdots s_\ell$ and some bracketing of $s_{\ell+1} \cdots s_k$, for some ℓ with $1 \leqslant \ell < k$. Now consider two cases:

- Suppose $\ell = 1$. By the assumption, the result of inserting brackets into $s_{\ell+1} \cdots s_k = s_2 \cdots s_k$ is equal to $s_2(s_3(\cdots s_k)\cdots)$. Thus $t = s_1\big(s_2(s_3(\cdots s_k)\cdots)\big)$, which is the result with $n = k$.
- Suppose $\ell > 1$. By the assumption, the result of the bracketing of $s_1 \cdots s_\ell$ is $s_1(s_2(\cdots s_\ell)\cdots)$ and the result of the bracketing of $s_{\ell+1} \cdots s_k$ is $s_{\ell+1}(s_{\ell+2}(\cdots s_k)\cdots)$. Thus

$$\begin{aligned} t &= \big(s_1(s_2(\cdots s_\ell)\cdots)\big)\big(s_{\ell+1}(s_{\ell+2}(\cdots s_k)\cdots)\big) \\ &= s_1\Big(\big((s_2(\cdots s_\ell)\cdots)\big)\big(s_{\ell+1}(s_{\ell+2}(\cdots s_k)\cdots)\big)\Big) && \text{[by associativity]} \\ &= s_1(s_2(s_3 \cdots s_k)\cdots), && \text{[by assumption with } n = k-1\text{]} \end{aligned}$$

  which is the result with $n = k$.

Hence, by induction, the result holds for all n. 1.1

Thus, by Proposition 1.1, there is no ambiguity in writing a product $s_1 s_2 \cdots s_n$ (where each $s_i \in S$): the product is the same regardless of how we insert the brackets.

Any group is also a semigroup. The most familiar example of a semigroup that is not a group is the set of natural numbers $\mathbb{N} = \{1, 2, 3, \ldots\}$

under the operation of addition. This is not a group since it does not contain inverses.


Note that, for our purposes, the set of natural numbers $\mathbb{N} = \{1, 2, 3, \ldots\}$ does not include 0.

Identity, monoid

Let e be an element of S. If $ex = x$ for all $x \in S$, the element e is a *left identity*. If $xe = x$ for all $x \in S$, the element e is a *right identity*. If $ex = xe = x$ for all $x \in S$, then e is a *two-sided identity* or simply an *identity*. A semigroup that contains an identity is called a *monoid*.

Zero

Let z be an element of S. If $zx = z$ for all $x \in S$, the element z is a *left zero*. If $xz = z$ for all $x \in S$, the element z is a *right zero*. If $zx = xz = z$ for all $x \in S$, then z is a *two-sided zero* or simply a *zero*.

EXAMPLE 1.2. Let us give some examples of semigroups:

a) The integers $\mathbb{Z}$ form a semigroup under two different operations: addition + and multiplication $\cdot$. The semigroup $(\mathbb{Z}, +)$ is a monoid with identity 0; but in $(\mathbb{Z}, \cdot)$, the element 0 is a zero.

Trivial semigroup

b) The *trivial semigroup* contains only one element e, with multiplication obviously defined by $ee = e$. Since e is (trivially) an identity, this semigroup is also called the *trivial monoid*.

Null semigroup

c) A *null semigroup* is a semigroup with a zero z in which the product of any two elements is z. It is easy to see that this multiplication is associative. Notice that we can define a null semigroup on any non-empty set by choosing some element z and defining all products to be z.

Right/left zero semigroup

d) If every element of S is a left zero (that is, $xy = x$ for all $x, y \in S$), then S is a *left zero semigroup*. If every element of S is a right zero (that is, $xy = y$ for all $x, y \in S$), then S is a *right zero semigroup*. We can define a left zero semigroup on any non-empty set X by defining the multiplication $xy = x$ for all $x, y \in X$; it is easy to see that this multiplication is associative. Similarly, we can define a right zero semigroup on any non-empty set X by defining the multiplication $xy = y$ for all $x, y \in X$.

e) Any ring is a semigroup under multiplication.

Uniqueness of an identity

PROPOSITION 1.3. *If e is a left identity of S and e' is a right identity of S then $e = e'$. Consequently, a semigroup contains at most one identity.*

Proof of 1.3. Since e is a left identity $ee' = e'$. Since e' is a right identity, $e = ee'$. Hence $e = ee' = e'$. 1.3

Uniqueness of a zero

PROPOSITION 1.4. *If z is a left zero of S and z' is a right zero of S then $z = z'$. Consequently, a semigroup contains at most one zero.*

Proof of 1.4. Since z is a left zero, $zz' = z$. Since z' is a right zero, $zz' = z'$. Hence $z = zz' = z'$. 1.4

It therefore makes sense to use the special notations 0 and 1 for the unique zero and identity of a semigroup. If we need to specify the zero or identity of a particular semigroup S, we will use 0_S and 1_S.

Adjoining an identity or zero

Let 1 be a new element not in the semigroup S. Extend the multiplication on S to $S \cup \{1\}$ by $1x = x1 = x$ for all $x \in S$ and $11 = 1$. It is easy to prove that this extended multiplication is associative. Then $S \cup \{1\}$ is a monoid with identity 1. Similarly, let 0 be a new element not in S and extend the multiplication on S to $S \cup \{0\}$ by $0x = x0 = 00 = 0$ for all $x \in S$. Again, this extended multiplication is associative. Then $S \cup \{0\}$ is a semigroup with zero 0. For any semigroup S, define

$$S^1 = \begin{cases} S & \text{if } S \text{ has an identity,} \\ S \cup \{1\} & \text{otherwise;} \end{cases}$$

$$S^0 = \begin{cases} S & \text{if } S \text{ has a zero,} \\ S \cup \{0\} & \text{otherwise.} \end{cases}$$

The semigroups S^1 and S^0 are called, respectively, the *monoid obtained by adjoining an identity to S if necessary* and the *semigroup obtained by adjoining a zero to S if necessary*.

Notation for maps

Throughout these notes, maps are written *on the right* and composed *left to right*. To clarify: let $\varphi : X \to Y$ and $\psi : Y \to Z$ be maps. The result of applying φ to an element x of X is denoted $x\varphi$. The composition of φ and ψ is denoted $\varphi \circ \psi$ or simply $\varphi\psi$, and is a map from X to Z with $x(\varphi\psi) = (x\varphi)\psi$ for all $x \in X$. For $X' \subseteq X$, the restriction of the map φ to X' is denoted $\varphi|_{X'}$.

Cartesian product

Let $\mathcal{X} = \{ X_i : i \in I \}$ (where I is an index set) be a collection of sets. Informally, the *cartesian product* $\prod_{i \in I} X_i$ of the sets in $\mathcal{X}$ is the set of tuples with $|I|$ components, where, for each $i \in I$, the i-th component is an element of X_i. More formally, the cartesian product $\prod_{i \in I} X_i$ is the set of maps σ from I to $\bigcup_{i \in I} X_i$ such that $i\sigma \in X_i$ for each $i \in I$. We think of the map σ as a tuple with i-th component $i\sigma$. We will use both map notation and (especially when the index set I is finite) tuple notation for elements of cartesian products. When $I = \{1, \ldots, n\}$ is finite, we write

Finitary cartesian product

$X_1 \times \ldots \times X_n$ for $\prod_{i \in I} X_i$, and we say the cartesian product is *finitary*. When the sets X_i are all equal to a set X (that is, when we consider the cartesian product of $|I|$ copies of the set X), we write X^I for $\prod_{i \in I} X_i$.

Direct product

Let $\mathcal{S} = \{ S_i : i \in I \}$ (where I is an index set) be a collection of semigroups. The *direct product* of the semigroups in $\mathcal{S}$ is their cartesian product $\prod_{i \in I} S_i$ with componentwise multiplication: $(i)(st) = (i)s(i)t$, or, using tuple notation,

$$(\ldots, s_i, \ldots)(\ldots, t_i, \ldots) = (\ldots, s_i t_i, \ldots).$$

It is easy to prove that this componentwise multiplication is associative, and so the direct product is itself a semigroup.

For $x \in S$ and $n \in \mathbb{N}$, define

Exponent

$$x^n = \overbrace{xx\cdots x}^{n \text{ times}}. \tag{1.1}$$

Notice that, in general, x^n is only defined for positive n. If S is a monoid, define $x^0 = 1_S$. Any element x^n, where $n \in \mathbb{N} \cup \{0\}$ is a *power* of x; if $n > 0$, it is a *positive power* of x. Notice that if S is not a monoid, then every power is a positive power.

Power, positive power

As an immediate consequence of Proposition 1.1, $x^m x^n = x^{m+n}$ for all $x \in S$ and $m, n \in \mathbb{N}$. If S is a monoid, then $x^m x^n = x^{m+n}$ for all $x \in S$ and $m, n \in \mathbb{N} \cup \{0\}$.

Exponent laws

Let $x \in S$ and consider the positive powers $x, x^2, x^3, \ldots$. There are two possibilities: either all these positive powers of x are distinct or there is some $k, \ell \in \mathbb{N}$ with $k < \ell$ such that $x^k = x^\ell$. In the latter case, x is said to be *periodic*; notice that in a finite semigroup, this latter case must hold. Choose ℓ as small as possible; then x^ℓ is the first positive power of x that is equal to some earlier positive power. Let $m = \ell - k$; then $x^k = x^{k+m}$. Repeatedly multiplying this equality by x^m, one sees that $x^k = x^{k+rm}$ for all $r \in \mathbb{N} \cup \{0\}$. Let $n \in \mathbb{N} \cup \{0\}$. Then $n = mr + i$ for some $r \in \mathbb{N} \cup \{0\}$ and $i \in \{0, \ldots, m-1\}$, and so $x^{k+n} = x^{k+mr+i} = x^{k+i}$. Therefore every power of x after x^k is equal to one of

Periodic element, index, period

$$x^k, x^{k+1}, \ldots, x^{k+m-1}.$$

Thus, by the minimality of the choice of ℓ, there are $k + m - 1$ distinct positive powers of x, namely

$$x, x^2, \ldots, x^{k-1}, x^k, x^{k+1}, \ldots, x^{k+m-1}.$$

We call k the *index* of x and m the *period* of x. A *periodic* semigroup is one in which every element is periodic. Note that all finite semigroups are periodic.

Periodic semigroup

An element x of S is an *idempotent* if $x^2 = x$. The set of idempotents of S is denoted $E(S)$. If every element of S is an idempotent, then S is a *semigroup of idempotents*. For example, a right zero semigroup is a semigroup of idempotents.

Idempotent, $E(S)$, semigroup of idempotents

For any subsets X and Y of S, define $XY = \{xy : x \in X, y \in Y\}$. Write xY for $\{x\}Y$ and Xy for $X\{y\}$. Since multiplication in S is associative, so is this product of subsets: for subsets X, Y, and Z of S, we have $X(YZ) = (XY)Z$. By analogy with (1.1), for $X \subseteq S$ and $n \in \mathbb{N}$, define

Product of subsets

$$X^n = \overbrace{XX\cdots X}^{n \text{ times}}.$$

The semigroup S is *nilpotent* if it contains a zero and there exists some $n \in \mathbb{N}$ such that $S^n = \{0\}$. The semigroup S is a *nilsemigroup* if it contains a zero and for every $x \in S$, there exists some $n \in \mathbb{N}$ such that $x^n = 0$.

Nilpotent semigroup, nilsemigroup

Note that this is incompatible with the definition of a 'nilpotent group'. A non-trivial group is never nilpotent in this semigroup sense.

Cancellativity

The semigroup S is *left-cancellative* if

$$(\forall x, y, z \in S)(zx = zy \Rightarrow x = y);$$

right-cancellative if

$$(\forall x, y, z \in S)(xz = yz \Rightarrow x = y);$$

and *cancellative* if it is both left- and right-cancellative. Note that a non-trivial semigroup with zero cannot be cancellative.

Commutativity

The semigroup S is *commutative* if $xy = yx$ for all $x, y \in S$. For instance, $(\mathbb{Z}, +)$ and $(\mathbb{Z}, \cdot)$ are both commutative. A non-trivial left zero semigroup is not commutative.

Left and right inverse

Let M be a monoid. Let $x \in M$. Suppose that there exists an element x' such that $xx' = 1$. Then x' is a *right inverse* for x, and x is *right invertible*. Similarly, suppose there exists an element x'' such that $x''x = 1$. Then x'' is a *left inverse* for x, and x is *left invertible*. If x is both left and right invertible, then x is *invertible*.

Right and left inverses coincide

PROPOSITION 1.5. *Let M be a monoid, and let $x \in M$. Suppose x is invertible and let x' be a right inverse of x and x'' a left inverse. Then $x' = x''$.*

Proposition 1.5 says that right and left inverses coincide *when they both exist*. The existence of one does not imply the existence of the other.

Proof of 1.5. Since x' and x'' are, respectively, right and left inverses of x, we have $xx' = 1$ and $x''x = 1$. Hence $x' = 1x' = x''xx' = x''1 = x''$. 1.5

Group

Thus if x is an invertible element of a monoid M, denote the unique right and left inverse of x by x^{-1}. A monoid in which every element is invertible is of course a *group*.

Regular element

Let $x \in S$. If there is an element $y \in S$ such that $xyx = x$, then the element x is *regular*. Notice that in this case, xy and yx are idempotent, since $(xy)^2 = xyxy = (xyx)y = xy$ and $(yx)^2 = yxyx = y(xyx) = yx$.

Regular semigroup

If every element of S is regular, then S is a *regular semigroup*.

Inverse

An element $x' \in S$ such that $x = xx'x$ and $x'xx' = x'$ is an *inverse* of x.

Notice that this is entirely different from the notion of left/right inverses above. We will never use 'inverse' (on its own) to refer to a left or right inverse.

PROPOSITION 1.6. *Let $x \in S$. Then x has an inverse if and only if x is regular.*

Proof of 1.6. Obviously if x has an inverse, then it is regular. So suppose x is regular. Then there exists $y \in S$ such that $xyx = x$. Let $x' = yxy$. Then $xx'x = x(yxy)x = (xyx)yx = xyx = x$ and $x'xx' = (yxy)x(yxy) = y(xyx)yxy = yxyxy = y(xyx)y = yxy = x'$, so x' is an inverse of x. 1.6


In the proof of Proposition 1.6, the element y might not be an inverse of x: for example, let S be a semigroup with a zero and let $x = 0$ and $y \neq 0$. Then $xyx = x$ but $yxy \neq y$.

An element x can have more than one inverse; see Example 1.7(e). The set of inverses of x is denoted $V(x)$. Notice also that a zero 0 of a semigroup has an inverse, namely 0 itself. In general, if $e \in S$ is idempotent, then $e^3 = e^2 = e$ and so e is an inverse of itself. In particular, every idempotent is regular.

Set of inverses $V(x)$

EXAMPLE 1.7. a) Let $U = \{0, \ldots, k\}$ for some $k \geqslant 0$. Define an operation $\triangle$ on U by $m \triangle n = \min\{m, n\}$. It is easy to see that $\triangle$ is associative, and so $(U, \triangle)$ is a semigroup. Notice that $0 \triangle m = m \triangle 0 = 0$ and $k \triangle m = m \triangle k = m$ for all $m \in U$. Hence U has zero 0 and identity k. Furthermore, $m \triangle m = m$ for all $m \in U$, so every element of U is an idempotent. Finally, $m \triangle n = n \triangle m$ for all $m, n \in U$ and so U is commutative.

b) Similarly, define an associative operation $\triangle$ on $\mathbb{N} \cup \{0\} = \{0, 1, 2, \ldots\}$ by $m \triangle n = \min\{m, n\}$. Then $(\mathbb{N} \cup \{0\}, \triangle)$ has a zero 0 but has no identity. It is commutative and all its elements are idempotents.

c) Consider the set of all 2×2 integer matrices:

$$M_2(\mathbb{Z}) = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} : a, b, c, d \in \mathbb{Z} \right\}.$$

With the usual matrix multiplication, $M_2(\mathbb{Z})$ is a monoid with identity $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and zero $\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$. It is easy to see that $M_2(\mathbb{Z})$ is not commutative, and that not all of its elements are idempotent. Since $M_2(\mathbb{Z})$ contains a zero, it is not cancellative.

d) Now let V be the set of all 2×2 integer matrices with non-zero determinant:

$$V = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} : a, b, c, d \in \mathbb{Z} \wedge \det \begin{bmatrix} a & b \\ c & d \end{bmatrix} \neq 0 \right\}.$$

Again, V is a monoid. Let $P, Q, R \in V$. Suppose $RP = RQ$. Since $\det R \neq 0$, the matrix R has a (left and right) inverse $R^{-1} \in M_2(\mathbb{Q})$. [Note that $R^{-1} \notin V$ whenever $\det R \neq \pm 1$, so V is not a group.] So $R^{-1}RP = R^{-1}RQ$ and so $P = Q$. Hence V is left-cancellative. Similarly, it is right-cancellative and therefore cancellative.

Rectangular band

e) Let L be a left zero semigroup and R a right zero semigroup. Let $B = L \times R$. The semigroup B is an $|L| \times |R|$ *rectangular band*, or simply a *rectangular band*. For $(\ell_1, r_1), (\ell_2, r_2) \in B$, we have

$$(\ell_1, r_1)(\ell_2, r_2) = (\ell_1 \ell_2, r_1 r_2) = (\ell_1, r_2),$$

since (in particular) ℓ_1 is a left zero and r_2 is a right zero. Thus every element of B is idempotent, since $(\ell, r)(\ell, r) = (\ell, r)$ for all $(\ell, r) \in B$. Furthermore, for any $(\ell_1, r_1), (\ell_2, r_2) \in B$, we have

$$(\ell_1, r_1)(\ell_2, r_2)(\ell_1, r_1) = (\ell_1 \ell_2 \ell_1, r_1 r_2 r_1) = (\ell_1, r_1)$$
$$(\ell_2, r_2)(\ell_1, r_1)(\ell_2, r_2) = (\ell_2 \ell_1 \ell_2, r_2 r_1 r_2) = (\ell_2, r_2).$$

Hence (ℓ_2, r_2) is an inverse of (ℓ_1, r_1). Thus every element is an inverse of every element.

The name 'rectangular band' comes from the following diagrammatic interpretation of the multiplication. The elements of the semigroup correspond to the cells of a grid whose rows are indexed by L and whose columns are indexed by R. So (ℓ_1, r_1) corresponds to the cell in row ℓ_1 and column r_1. In terms of cells, the product of two elements is the cell in the row of the first multiplicand and the column of the second multiplicand; see Figure 1.1. [The reason for indexing rows by the first coordinate and columns by the second coordinate will become clear in Chapter 4.]

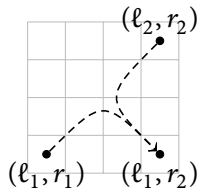


FIGURE 1.1
Diagrammatic interpretation of multiplication in a rectangular band.

Opposite semigroup

The *opposite semigroup* S^{opp} of S is the semigroup with the same set as S but 'reversed multiplication'. That is, for $x, y \in S$, the product xy in S^{opp} is equal to the product yx in S. It is easy to check that S^{opp} is indeed a semigroup. Notice that if S is commutative, then S^{opp} and S are the same semigroup.

GENERATORS AND SUBSEMIGROUPS

Subsemigroup

A non-empty subset T of S is a *subsemigroup* if it is closed under multiplication; that is, if $TT \subseteq T$. A *proper* subsemigroup is any subsemigroup except S itself. A *submonoid* is a subsemigroup that happens to be a monoid. A *subgroup* is a subsemigroup that happens to be a group.

PROPOSITION 1.8. *The set of invertible elements of a monoid forms a subgroup.*

Proof of 1.8. Let T be the set of invertible elements of a monoid M. Note that T is non-empty since $1 \in T$. Let $x, y \in T$. Then since x and y are invertible we have $y^{-1}x^{-1}xy = y^{-1}y = 1$ and $xyy^{-1}x^{-1} = xx^{-1} = 1$.

Hence xy is invertible and so lies in T. Hence T is a subsemigroup of M. Furthermore, $1 \in T$ is also an identity for T and so T is a submonoid of M. Finally, let $x \in T$; then $xx^{-1} = x^{-1}x = 1$ and so x^{-1} is invertible and thus $x^{-1} \in T$. Hence T is closed under taking inverses. Therefore T is a subgroup of M. 1.8

The set of invertible elements of a monoid is called its *group of units*; Proposition 1.8 justifies this name. Group of units

The following result is a useful characterization of subgroups of a semigroup:

LEMMA 1.9. *Let G be a non-empty subset of S. Then $gG = Gg = G$ for all $g \in G$ if and only if G is a subgroup of S.*

Proof of 1.9. Notice first that if G is a subgroup and $g \in G$, then $G = gg^{-1}G \subseteq gG \subseteq G$, so $G = gG$. Similarly, $G = Gg$.

For the converse, suppose that $gG = Gg = G$ for all $g \in G$. For any $g, h \in G$, the product gh lies in $gG = G$. Hence G is a subsemigroup.

Let $g \in G$. Since $G = Gg$, it follows that $g \in Gg$, and so there exists $e \in G$ such that $g = eg$. Let $h \in G$. Since $G = gG$, there exists $x \in G$ such that $h = gx$. Hence $eh = egx = gx = h$. Since $h \in G$ was arbitrary, e is a left identity for G. Similarly G contains a right identity f, and so $e = f$ is an identity for G by Proposition 1.3. So G is a submonoid with identity 1_G.

Finally, since $1_G \in gG = Gg$, the element g is right and left invertible and its right and left inverses coincide by Proposition 1.5. Since $g \in G$ was arbitrary, G is a subgroup. 1.9

Let T be a non-empty subset of S. The subset T is a *left ideal* of S if it is closed under left multiplication by any element of S; that is, if $ST \subseteq T$. It is a *right ideal* of S if it is closed under right multiplication by any element of S; that is, if $TS \subseteq T$. It is a *two-sided ideal*, or simply an *ideal*, of S if it is closed under both left and right multiplication by elements of S; that is, if $ST \cup TS \subseteq T$. Every ideal, whether left, right, or two-sided, is a subsemigroup. Ideal

For any $x \in S$, define Principal ideal

$$\begin{aligned} L(x) &= S^1x = \{x\} \cup Sx, \\ R(x) &= xS^1 = \{x\} \cup xS, \\ J(x) &= S^1xS^1 = \{x\} \cup xS \cup Sx \cup SxS. \end{aligned}$$

Then $L(x)$, $R(x)$, and $J(x)$ are, respectively, the *principal left ideal generated by x*, the *principal right ideal generated by x*, and the *principal ideal generated by x*. As their names imply, they are, respectively, a left ideal, a right ideal, and a (two-sided) ideal.

EXAMPLE 1.10. a) Consider the semigroup $(\mathbb{N}, +)$. Let $n \in \mathbb{N}$ and let $I_n = \{m \in \mathbb{N} : m \geqslant n\}$. Then I_n is an ideal of $\mathbb{N}$; indeed, $I_n = L(n) = R(n) = J(n)$.

b) Let S be a right zero semigroup. Let T be a non-empty subset of S. Then $ST = T$ since $xy = y$ for any $x \in S$ and $y \in T$. So T is a left ideal of S. On the other hand, $TS = S$ and so T is a right ideal if and only if $T = S$.

c) Let G be a group. Let T be a non-empty subset of G. For any $x \in G$ and $y \in T$, we have $x = xy^{-1}y \in Gy$; hence $Gy = G$. So T is a left ideal if and only if $T = G$; similarly T is a right ideal if and only if $T = G$. So the only left ideal or right ideal of G is G itself.

Generating a subsemigroup

Let $\mathcal{T} = \{T_i : i \in I\}$ be a collection of subsemigroups of S. It is easy to see that if their intersection $\bigcap \mathcal{T} = \bigcap_{i \in I} T_i$ is non-empty, it is also a subsemigroup. So let X be a non-empty subset of S and let $\mathcal{T}$ be the collection of subsemigroups of S that contain X. The collection $\mathcal{T}$ has at least one member, namely the semigroup S itself, and every subsemigroup in $\mathcal{T}$ contains X, so $\bigcap \mathcal{T}$ is non-empty and is thus a subsemigroup. Indeed, it is the smallest subsemigroup of S that contains X. This subsemigroup, denoted $\langle X \rangle$, is called the *subsemigroup generated by* X.

Generating set

If $X \subseteq S$ is such that $\langle X \rangle = S$, then X is a *generating set* for S and X *generates* S. If there is a finite generating set for S, then S is said to be *finitely generated*.

PROPOSITION 1.11. *Let X be a non-empty subset of S. Then $\langle X \rangle = \{x_1 x_2 \cdots x_n : n \in \mathbb{N}, x_i \in X\}$.*

Proof of 1.11. Let $U = \{x_1 x_2 \cdots x_n : n \in \mathbb{N}, x_i \in X\}$. Then U is closed under multiplication and so is a subsemigroup of S. Furthermore, $X \subseteq U$. Hence U must be one of the T_i in $\mathcal{T}$, and so $\langle X \rangle \subseteq U$. Since $X \subseteq \langle X \rangle$ and $\langle X \rangle$ is closed under multiplication, $U \subseteq \langle X \rangle$. Therefore $\langle X \rangle = U$. 1.11

Monogenic semigroup

Suppose S is generated by a single element x; that is, $S = \langle \{x\} \rangle$ (which we abbreviate to $S = \langle x \rangle$). Then S is a *monogenic semigroup*, and, by Proposition 1.11, $S = \{x^n : n \in \mathbb{N}\}$. If the element x is periodic with index k and period m, then $S = \{x, x^2, \ldots, x^{k+m-1}\}$. Let

$$K = \{x^k, x^{k+1}, \ldots, x^{k+m-1}\}.$$

It is easy to see that K is an ideal of S.

PROPOSITION 1.12. *The ideal K is a subgroup of S.*

Proof of 1.12. Let $I = \{k, k+1, \ldots, k+m-1\}$, so that $K = \{x^n : n \in I\}$. Then I is a complete set of representatives for congruence classes of the integers modulo m. In particular there is some $p \in I$ such that $p \equiv 0 \bmod m$; note that $p = rm$ for some $r \in \mathbb{N}$. Let $e = x^p = x^{rm}$. Then

$x^n e = x^n x^{rm} = x^{n+rm} = x^n$ and similarly $ex^n = x^n$ for any $n \in I$; hence e is an identity for K.

Now let $n \in I$. Choose $q \in I$ with $q \equiv -n \bmod m$. Then $q + n \equiv 0 \bmod m$ and so $q + n = sm$ for some $s \in \mathbb{N}$. Hence $x^q x^n = x^{q+n} = x^{sm}$. Since $sm \geqslant k$ and since rm is the unique multiple of m in I, it follows that $sm = rm + tm$ for some $t \in \mathbb{N} \cup \{0\}$. Hence $x^q x^n = x^{rm+tm} = x^{rm} = e$, and similarly $x^n x^q = e$. Hence x^q is a right and left inverse for x^n; since $n \in I$ was arbitrary, every element of K has an inverse in K. 1.12

Note that x^m may not be an identity for m. It is true that $x^n x^m = x^m x^n = x^n$, but if $k > m$, then $x^m \notin K$.

Generating a submonoid

Given a subset X of a monoid M, we can also define the submonoid generated by X. Let $\mathcal{T}$ be the collection of submonoids of M that contain $X \cup \{1_M\}$. The intersection of the submonoids in $\mathcal{T}$ is non-empty and thus a submonoid. This is the smallest submonoid of M with identity 1_M that contains X. This submonoid, denoted $\mathrm{Mon}\langle X\rangle$, is called the *submonoid generated by* X. Reasoning similar to the proof of Proposition 1.11 yields the following result:

PROPOSITION 1.13. *Let* $X \subseteq M$. *Then* $\mathrm{Mon}\langle X\rangle = \{\, 1_M x_1 x_2 \cdots x_n : n \in \mathbb{N} \cup \{0\}, x_i \in X \,\}$. 1.13

Monoid generator

Essentially, when we generate a submonoid of a monoid, we always include the identity of the monoid. If $X \subseteq M$ is such that $\mathrm{Mon}\langle X\rangle = M$, then X is a *monoid generating set* for M and X *generates* M *as a monoid*.

Generating set and monoid generating set

Notice that if X is a generating set for M, then X is also a monoid generating set; on the other hand, if X is a monoid generating set for M, then $X \cup \{1_M\}$ is a generating set for M. Thus M is finitely generated if and only if there is a finite monoid generating set for M.

BINARY RELATIONS

Recall that a relation ρ between a set X and a set Y is simply a subset of $X \times Y$, and $x \;\rho\; y$ is equivalent to $(x, y) \in \rho$. The *identity relation* on X is the relation

Identity relation

$$\mathrm{id}_X = \{\, (x, x) : x \in X \,\}.$$

The *converse* ρ^{-1} of ρ is the relation

Converse of a relation

$$\rho^{-1} = \{\, (y, x) : (x, y) \in \rho \,\}.$$

The converse relation ρ^{-1} is not in general a left or right inverse of ρ, even when ρ is a map.

Composition of relations

Let ρ be a relation between X and Y and σ be a relation between Y

and Z. Define the composition of ρ and σ to be

$$\rho \circ \sigma = \{ (x,z) \in X \times Z : (\exists y \in Y)\big((x\ \rho\ y) \wedge (y\ \sigma\ z)\big) \}. \tag{1.2}$$

Notice that $\rho \circ \sigma$ is a relation between X and Z. Furthermore, notice that $\rho \circ \mathrm{id}_Y = \rho$ and $\mathrm{id}_Y \circ \sigma = \sigma$.

Partial/full map

For any $x \in X$, let $x\rho = \{ y \in Y : x\ \rho\ y \}$. Then ρ is a *partial map* from X to Y if $|x\rho| \leqslant 1$ for all $x \in X$. Furthermore, ρ is a *full map*, or simply a *map* from X to Y if $|x\rho| = 1$ for all $x \in X$.

Suppose ρ is a partial map from X to Y. When $x\rho$ is the empty set, we say that $x\rho$ is undefined; when $x\rho$ is the singleton set $\{y\}$, we say that $x\rho$ is defined and write $x\rho = y$ instead of $x\rho = \{y\}$.

The definition of a map given here, and the notation in the last paragraph, agree with the standard concept and notation of a map. Furthermore, when ρ and σ are maps, (1.2) simply defines the usual composition of maps. Thus we have recovered the usual notion of maps in a more general setting.

Domain, image, preimage

For any partial map ρ from X to Y, the *domain* of ρ is the set

$$\mathrm{dom}\,\rho = \{ x \in X : (\exists y \in Y)\big((x,y) \in \rho\big) \}. \tag{1.3}$$

That is, $\mathrm{dom}\,\rho$ is the subset of X on which ρ is defined. If ρ is a map, we have $\mathrm{dom}\,\rho = X$. The *image* of ρ is the set

$$\mathrm{im}\,\rho = \{ y \in Y : (\exists x \in X)\big((x,y) \in \rho\big) \}. \tag{1.4}$$

The *preimage* under ρ of $Y' \subseteq Y$ is the set

$$\begin{aligned} Y'\rho^{-1} &= \{ x \in X : (\exists y \in Y')((y,x) \in \rho^{-1}) \} \\ &= \{ x \in X : (\exists y \in Y')((x,y) \in \rho) \}. \end{aligned}$$

Binary relations, $\mathcal{B}_X$

We will be particularly interested in binary relations on X; that is, relations from X to itself. Let $\mathcal{B}_X$ denote the set of all binary relations on X. It is easy to show that $\circ$ is an associative operation on $\mathcal{B}_X$ and so $(\mathcal{B}_X, \circ)$ is a semigroup, called the *semigroup of binary relations on* X. Furthermore, id_X is an identity and so $\mathcal{B}_X$ is a monoid.

Partial/full transformation

$\mathcal{P}_X$, $\mathcal{T}_X$, $\mathcal{S}_X$

A partial map from X to itself is called a *partial transformation* of X. A map from X to itself is called a *full transformation*, or simply a *transformation* of X. The set of all partial transformations of X is $\mathcal{P}_X$; the set of all [full] transformations of X is $\mathcal{T}_X$. Finally, $\mathcal{S}_X$ denotes the set of bijections on X. This is the well-known *symmetric group* on X. Clearly $\mathcal{S}_X \subseteq \mathcal{T}_X \subseteq \mathcal{P}_X \subseteq \mathcal{B}_X$.

PROPOSITION 1.14. a) *$\mathcal{P}_X$ is a submonoid of $\mathcal{B}_X$;*

b) *$\mathcal{T}_X$ is a submonoid of $\mathcal{P}_X$;*

c) *$\mathcal{S}_X$ is a subgroup of $\mathcal{T}_X$.*

Proof of 1.14. a) Let $\rho, \sigma \in \mathcal{P}_X$ and suppose $y, y' \in x(\rho \circ \sigma)$. Then by the definition of $\circ$, there exist $z, z' \in X$ such that $(x, z) \in \rho$ and $(z, y) \in \sigma$, and $(x, z') \in \rho$ and $(z', y') \in \sigma$. Since $\rho \in \mathcal{P}_X$, we have $|x\rho| \leqslant 1$ and so $z = z'$. Since $\sigma \in \mathcal{P}_X$, we have $|z\sigma| \leqslant 1$ and so $y = y'$. Hence $|x(\rho \circ \sigma)| \leqslant 1$ and so $\rho \circ \sigma \in \mathcal{P}_X$.

b) Let $\rho, \sigma \in \mathcal{T}_X$. Let $x \in X$. Since $\rho \in \mathcal{T}_X$, we have $|x\rho| = 1$. So let $z = x\rho$. Since $\sigma \in \mathcal{T}_X$, we have $|z\sigma| = 1$. So $x(\rho \circ \sigma)$ contains $(x, z\sigma)$ and so $|x(\rho \circ \sigma)| \geqslant 1$. By part a), $|x(\rho \circ \sigma)| = 1$. Therefore $\rho \circ \sigma \in \mathcal{T}_X$.

c) This is immediate because the composition of two bijections is a bijection. 1.14

In light of Proposition 1.14, $\mathcal{T}_X$ is called the *semigroup of transformations on* X and $\mathcal{P}_X$ is called the *semigroup of partial transformations on* X.

Two-line notation for transformations

Any bijection $\rho \in \mathcal{S}_X$ can be denoted by the usual disjoint cycle notation from group theory. A partial (or full) transformation $\rho \in \mathcal{P}_X$ can be denoted using a $2 \times |X|$ matrix: the $(1, x)$-th entry is x and the $(2, x)$-th entry is either $x\rho$ (when $x\rho$ is defined) or $*$ (indicating that $x\rho$ is undefined). For example, if $X = \{1, 2, 3\}$ and $1\rho = 2$, and 2ρ is undefined, and $3\rho = 1$, then

$$\rho = \begin{pmatrix} 1 & 2 & 3 \\ 2 & * & 1 \end{pmatrix}.$$

EXAMPLE 1.15. Let $X = \{1, 2\}$. Then

- $\mathcal{S}_X$ consists of two elements:

$$\mathrm{id}_X = \begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix} \text{ and } \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix};$$

- $\mathcal{T}_X$ consists of four elements: the two elements in $\mathcal{S}_X$, and the transformations

$$\begin{pmatrix} 1 & 2 \\ 1 & 1 \end{pmatrix} \text{ and } \begin{pmatrix} 1 & 2 \\ 2 & 2 \end{pmatrix};$$

- $\mathcal{P}_X$ consists of nine elements: the four elements in $\mathcal{T}_X$, and the partial transformations

$$\begin{pmatrix} 1 & 2 \\ 1 & * \end{pmatrix}, \begin{pmatrix} 1 & 2 \\ 2 & * \end{pmatrix}, \begin{pmatrix} 1 & 2 \\ * & 1 \end{pmatrix}, \begin{pmatrix} 1 & 2 \\ * & 2 \end{pmatrix}, \text{ and } \begin{pmatrix} 1 & 2 \\ * & * \end{pmatrix};$$

- $\mathcal{B}_X$ consists of all sixteen possible subsets of $X \times X$, including the empty set $\varnothing$ and $X \times X$ itself.

Let us illustrate how elements of the semigroups of partial and full transformations multiply:

EXAMPLE 1.16. Let $X = \{1, 2, 3\}$.

Multiplication in $\mathcal{T}_X$

a) Let

$$\rho = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 1 \end{pmatrix} \text{ and } \sigma = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 2 & 3 \end{pmatrix}$$

be elements of $\mathcal{T}_X$. Let us compute the product $\rho\sigma$. First, ρ contains the pair $(1, 3)$ and σ contains the pair $(3, 3)$, so $\rho\sigma$ contains the pair $(1, 3)$. Using our notation for partial and full maps, this says that $1\rho = 3$ and $3\sigma = 3$, and thus $1\rho\sigma = 3\sigma = 3$. Similarly, $2\rho\sigma = 1\sigma = 2$ and $3\rho\sigma = 1\sigma = 2$. Hence

$$\rho\sigma = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 2 \end{pmatrix}.$$

Multiplication in $\mathcal{P}_X$

b) Let

$$\rho = \begin{pmatrix} 1 & 2 & 3 \\ 3 & * & 2 \end{pmatrix} \text{ and } \sigma = \begin{pmatrix} 1 & 2 & 3 \\ 1 & * & 2 \end{pmatrix}$$

be elements of $\mathcal{P}_X$. Let us compute the product $\rho\sigma$. First, $1\rho\sigma = 3\sigma = 2$; this part of the computation is just like the case of a full map. Next, 2ρ is undefined: that is, ρ does not contain the pair $(2, x)$ for any $x \in X$. Hence $\rho\sigma$ cannot contain the pair $(2, y)$ for any $y \in X$. That is, $2\rho\sigma$ is undefined. Finally, $3\rho = 2$, but σ does not contain the pair $(2, x)$ for any $x \in X$, and hence $\rho\sigma$ cannot contain the pair $(3, x)$ for any $x \in X$. That is, $3\rho\sigma$ is undefined. Hence

$$\rho\sigma = \begin{pmatrix} 1 & 2 & 3 \\ 2 & * & * \end{pmatrix}.$$

[During this computation, it may be helpful to think of '$*$' as an additional element of X that is mapped to itself by every partial transformation of $\mathcal{P}_X$. Then one can think 'ρ maps 2 to $*$ and σ maps $*$ to $*$, so $\rho\sigma$ maps 2 to $*$' and 'ρ maps 3 to 2 and σ maps 2 to $*$, so $\rho\sigma$ maps 3 to $*$'. Remember, however, that $*$ is not an element of X, but is simply a notational convenience to indicate where a partial transformation is undefined.]

Reflexive, (anti-)symmetric, transitive

There are several important properties that a binary relation may have: a relation $\rho \in \mathcal{B}_X$ is

- *reflexive* if $x \rho x$ for all $x \in X$, or, equivalently, if $\mathrm{id}_X \subseteq \rho$;
- *symmetric* if $x \rho y \Rightarrow y \rho x$ for all $x, y \in X$, or, equivalently, if $\rho = \rho^{-1}$;
- *anti-symmetric* if $(x \rho y) \wedge (y \rho x) \Rightarrow x = y$ for all $x, y \in X$, or, equivalently, if $\rho \cap \rho^{-1} \subseteq \mathrm{id}_X$;

- *transitive* if $(x \ \rho \ y) \wedge (y \ \rho \ z) \Rightarrow x \ \rho \ z$ for all $x, y, z \in X$, or, equivalently, if $\rho^2 \subseteq \rho$.

Notice that 'anti-symmetric' is not the same as 'not symmetric': for example, the identity relation id_X is both symmetric and anti-symmetric.

An *equivalence relation* is a relation that is reflexive, symmetric, and transitive. An equivalence relation on X partitions the set X into *equivalence classes*, each made up of related elements.

Equivalence relation

ORDERS AND LATTICES

Let $\rho \in \mathcal{B}_X$. The binary relation ρ is a *partial order* if it is reflexive, anti-symmetric, and transitive. We normally use symbols like $\leqslant$, $\preccurlyeq$, and $\sqsubseteq$ for partial orders. We write $x < y$ to mean that $x \leqslant y$ and $x \neq y$; the obvious analogies apply for symbols like $\prec$ and $\sqsubset$. A *partially ordered set* or *poset* is a set X equipped with a partial order $\leqslant$, formally denoted $(X, \leqslant)$.

Partial order

If $(X, \leqslant)$ is a partially ordered set and Y is a subset of X, then Y 'inherits' the partial order $\leqslant$ from X. That is, the restriction of the relation $\leqslant$ to Y (that is, $\leqslant \cap (Y \times Y)$) is a partial order on Y, and so Y is also a partially ordered set. We use the same notation for the original partial order on X and for its restriction to Y.

A *Hasse diagram* of a partial order $\leqslant$ on a set X is a diagrammatic representation of $\leqslant$. Every element of X is represented by a point on the plane, arranged so that x appears below y whenever $x < y$. If $x < y$ and there is no element z such that $x < z < y$, then a line segment is drawn between x and y.

Hasse diagram

Suppose $\leqslant$ is a partial order on X. Two elements $x, y \in X$ are *comparable* if $x \leqslant y$ or $y \leqslant x$. The partial order $\leqslant$ is a *total order*, or simply an *order*, if all pairs of elements of X are comparable.

Total order

Suppose $\leqslant$ is a partial order on X. A *chain* is a subset Y of X in which every pair of elements are comparable. An *antichain* is a subset Y of X in which no pair of distinct elements is comparable. Note that it is possible for X itself to be a chain or an antichain. If X is a chain (respectively, antichain), then any subset of X is also a chain (respectively, antichain).

Chain, antichain

EXAMPLE 1.17. a) For example, the relation $\leqslant$ on the integers $\mathbb{Z}$ is a partial order: it is reflexive, since $m \leqslant m$ for all m; it is anti-symmetric, since $m \leqslant n$ and $n \leqslant m$ imply $m = n$; and it is transitive, since $m \leqslant n$ and $n \leqslant p$ imply $m \leqslant p$.

b) Let X be a set. Recall that the power set $\mathbb{P}X$ is the set of all subsets of X. The relation $\subseteq$ on $\mathbb{P}X$ is a partial order. Figure 1.2 shows the Hasse diagram of $\mathbb{P}\{1, 2, 3\}$. Notice that $\subseteq$ is not a total order: for instance, $\{1\}$ and $\{2, 3\}$ are not comparable.

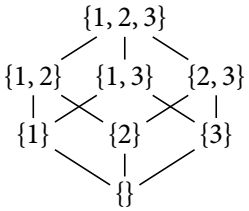


FIGURE 1.2
Hasse diagram for $\subseteq$ on $\mathbb{P}\{1, 2, 3\}$.

c) Let $|$ be the divisibility relation on $\mathbb{N}$; that is, $x \mid y$ if and only if there exists $p \in \mathbb{N}$ such that $y = px$. Then $|$ is reflexive, since $x \mid x$ for all $x \in \mathbb{N}$. It is anti-symmetric, since $x \mid y$ and $y \mid x$ imply $y = px$ and $x = p'y$ for some $p, p' \in \mathbb{N}$, which implies $x = p'px$ and so $p = p' = 1$, which implies $x = y$. It is transitive, since $x \mid y$ and $y \mid z$ imply $y = px$ and $z = p'y$ for some $p, p' \in \mathbb{N}$, which implies $z = (p'p)x$ and so $x \mid z$. So $|$ is a partial order on $\mathbb{N}$.

Minimal/minimum, maximal/maximum

If $x \in X$ is such that there is no element $y \in X$ with $y < x$ (respectively, $x < y$), then x is *minimal* (respectively, *maximal*). If $x \in X$ is such that for all elements $y \in X$, we have $x \leqslant y$ (respectively $y \leqslant x$), then x is a *minimum* (respectively, *maximum*). Therefore, in summary:

$$\begin{aligned} x \text{ is minimal} &\Leftrightarrow (\forall y \in X)(y \leqslant x \Rightarrow y = x);\\ x \text{ is minimum} &\Leftrightarrow (\forall y \in X)(x \leqslant y);\\ x \text{ is maximal} &\Leftrightarrow (\forall y \in X)(x \leqslant y \Rightarrow y = x);\\ x \text{ is maximum} &\Leftrightarrow (\forall y \in X)(y \leqslant x). \end{aligned}$$

Notice that a minimum element is also minimal, but that the converse does not hold. A poset does not have to contain minimum or minimal elements. It contains at most one minimum element, for if x_1 and x_2 are both minimum, then $x_1 \leqslant x_2$ and $x_2 \leqslant x_1$, and so $x_1 = x_2$ by anti-symmetry. It may contain many distinct minimal elements.

EXAMPLE 1.18. a) The poset $(\mathbb{Z}, \leqslant)$ does not contain either maximal elements or minimal elements.

b) Let $X = \{x, y_1, y_2\}$; define $\leqslant$ on X by

$$\begin{aligned} u &\leqslant u && \text{for all } u \in X,\\ y_1 &\leqslant x,\\ y_2 &\leqslant x. \end{aligned}$$

The Hasse diagram for $(X, \leqslant)$ is as shown in Figure 1.3(a): x is a (necessarily unique) maximum element, and y_1 and y_2 are both minimal elements.

c) Let $X = \{x, y, z_1, z_2, \ldots\}$ and define $\leqslant$ by

$$\begin{aligned} u &\leqslant u && \text{for all } u \in X,\\ y &\leqslant x,\\ z_i &\leqslant x && \text{for all } i \in \mathbb{N},\\ z_i &\leqslant z_j && \text{for all } i, j \in \mathbb{N} \text{ with } i \geqslant j. \end{aligned}$$

The Hasse diagram for $(X, \leqslant)$ is as shown in Figure 1.3(b): x is a (necessarily unique) maximum element, and y is the unique minimal element, but y is not a minimum.

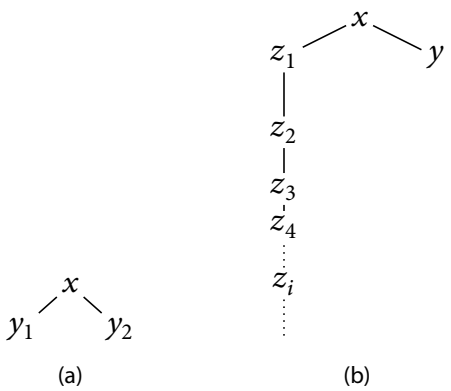


FIGURE 1.3
Examples of partial orders, illustrating minimal/maximal and minimum/maximum elements: (a) has a maximum x and two minimal elements y_1 and y_2; (b) has a unique minimal element y but has no minimum element.

Partial order of idempotents

There is a natural partial order of idempotents of a semigroup S that will re-appear in several different settings. Define the relation $\preccurlyeq$ on the set of idempotents $E(S)$ by $e \preccurlyeq f \Leftrightarrow ef = fe = e$.

PROPOSITION 1.19. *The relation $\preccurlyeq$ is a partial order.*

Proof of 1.19. Since $e^2 = e$, we have $e \preccurlyeq e$ and so $\preccurlyeq$ is reflexive. If $e \preccurlyeq f$ and $f \preccurlyeq e$, then $ef = fe = e$ and $fe = ef = f$ and so $e = f$; hence $\preccurlyeq$ is anti-symmetric. If $e \preccurlyeq f$ and $f \preccurlyeq g$, then $ef = fe = e$ and $fg = gf = f$ and so $ge = gfe = fe = e$ and $eg = efg = ef = e$ and thus $e \preccurlyeq g$; hence $\preccurlyeq$ is transitive. Therefore $\preccurlyeq$ is a partial order. 1.19

Lower bound

Let $\leqslant$ be a partial order on a set X. Let $Y \subseteq X$. A *lower bound* for Y is any element z of X such that $z \leqslant y$ for all $y \in Y$. Let B be the set of lower bounds for Y. If B is non-empty and has a maximum element z, then z is the *greatest lower bound* or *meet* or *infimum*of Y. The meet of Y, if it exists, is unique and is denoted by $\bigsqcap Y$, or, in the case where $Y = \{x, y\}$, by $x \sqcap y$.

Greatest lower bound, meet

Semilattice

If $x \sqcap y$ exists for all $x, y \in X$, then X is a *meet semilattice* or *lower semilattice*. If $\bigsqcap Y$ exists for all $Y \subseteq X$, then X is a *complete meet semilattice* or *complete lower semilattice*.

Upper bound, join

The obvious definitions apply for *upper bound*, *least upper bound* or *join* or *supremum*, $\bigsqcup Y$, $x \sqcup y$, *join semilattice* or *upper semilattice*, and *complete join semilattice* or *complete upper semilattice*.


Most texts use $\wedge$, $\vee$, $\bigwedge$, and $\bigvee$ in place of $\sqcap$, $\sqcup$, $\bigsqcap$, and $\bigsqcup$. The square variants are used here to avoid confusion with the symbols for logical conjunction ('and') $\wedge$ and disjunction ('or') $\vee$.

Lattice

The partially ordered set $(X, \leqslant)$ is a *lattice* if it is an upper and lower semilattice. It is a *complete lattice* if it is an complete upper semilattice and complete lower semilattice.

EXAMPLE 1.20. a) In the example of the relation $\subseteq$ on the power set $\mathbb{P}\{1, 2, 3\}$, we have $\{1, 2\} \sqcap \{1, 3\} = \{1\}$ and $\{1, 2\} \sqcap \{3\} = \{\}$. Indeed, $\big(\mathbb{P}\{1, 2, 3\}, \subseteq\big)$ is a complete lattice.

b) Let $X = \{t, x, y, z_1, z_2, \ldots\}$ and define $\leqslant$ by

$$\begin{aligned} x &\leqslant t, \\ y &\leqslant t, \\ z_i &\leqslant t && \text{for all } i \in \mathbb{N}, \\ z_i &\leqslant x && \text{for all } i \in \mathbb{N}, \\ z_i &\leqslant y && \text{for all } i \in \mathbb{N}, \\ z_i &\leqslant z_j && \text{for all } i, j \in \mathbb{N} \text{ with } i \leqslant j. \end{aligned}$$

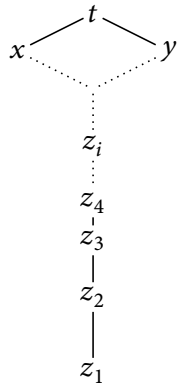


FIGURE 1.4
Partial Hasse diagram for the partially ordered set $(X, \leqslant)$.

Figure 1.4 shows a partial Hasse diagram for $(X, \leqslant)$. Notice that x and y do not have a meet, but that every pair of elements has a join. So $(X, \leqslant)$ is an upper semilattice but not a lower semilattice. However, it is not a *complete* upper semilattice because the subset $\{ z_i : i \in \mathbb{N} \}$ does not have a join.

Semilattice = commutative semigroup of idempotents

THEOREM 1.21. a) *Let $(X, \leqslant)$ be a non-empty lower semilattice. Then $(X, \sqcap)$ is a commutative semigroup of idempotents.*

Conversely, let $(S, \circ)$ be a commutative semigroup of idempotents. Define a relation $\leqslant$ on S by $x \leqslant y \Leftrightarrow x \circ y = x$. Then $\leqslant$ is a partial order and $(S, \leqslant)$ is a lower semilattice.

b) *Let $(X, \leqslant)$ be a non-empty upper semilattice. Then $(X, \sqcup)$ is a commutative semigroup of idempotents.*

Conversely, let $(S, \circ)$ be a commutative semigroup of idempotents. Define a relation $\leqslant$ on S by $x \leqslant y \Leftrightarrow x \circ y = y$. Then $\leqslant$ is a partial order and $(S, \leqslant)$ is an upper semilattice.

Proof of 1.21. We prove part a); the reasoning for part b) is dual. Suppose $(X, \leqslant)$ is a lower non-empty semilattice. Let $x, y, z \in X$. First, $x \sqcap (y \sqcap z)$ and $(x \sqcap y) \sqcap z$ are both the meet of $\{x, y, z\}$ and hence $x \sqcap (y \sqcap z) = (x \sqcap y) \sqcap z$. So $\sqcap$ is associative. Next, $x \sqcap y$ and $y \sqcap x$ are both the meet of $\{x, y\}$, and so $x \sqcap y = y \sqcap x$. So $(X, \sqcap)$ is commutative. The meet of $\{x\}$ is x itself, so $x \sqcap x = x$. Hence every element of $(X, \sqcap)$ is idempotent. So $(X, \sqcap)$ is a commutative semigroup of idempotents.

Suppose $(S, \circ)$ is a commutative semigroup of idempotents and define $\leqslant$ as in the statement of the result. Let $x, y, z \in S$. First, x is idempotent, and so $x \circ x = x$, and thus $x \leqslant x$. Hence $\leqslant$ is reflexive. Second, suppose that $x \leqslant y$ and $y \leqslant x$. Then $x \circ y = x$ and $y \circ x = y$. Since $(S, \circ)$ is commutative, this shows that $x = y$. Hence $\leqslant$ is anti-symmetric. Third, suppose $x \leqslant y$ and $y \leqslant z$. Then $x \circ y = x$ and $y \circ z = y$. So $x \circ z = (x \circ y) \circ z = x \circ (y \circ z) = x \circ y = x$, and so $x \leqslant z$. Hence $\leqslant$ is transitive.

Finally, we want to show that $x \sqcap y = x \circ y$. First of all $(x \circ y) \circ x = (x \circ y)$, so $x \circ y \leqslant x$ and similarly $x \circ y \leqslant y$. So $x \circ y$ is a lower bound for $\{x, y\}$. Let z be some lower bound for $\{x, y\}$. Then $z \leqslant x$ and $z \leqslant y$. Hence $z \circ x = z$ and $z \circ y = z$. So $z \circ (x \circ y) = (z \circ x) \circ y = z \circ y = z$, and so $z \leqslant (x \circ y)$. Hence $x \circ y$ is the greatest lower bound for $\{x, y\}$. Thus $(S, \leqslant)$ is a lower semilattice. 1.21

HOMOMORPHISMS

Homomorphism

Let S and T be semigroups. A map $\varphi : S \to T$ is a *homomorphism* if $(xy)\varphi = (x\varphi)(y\varphi)$ for all $x, y \in S$. If S and T are monoids, then φ is a *monoid homomorphism* if $(xy)\varphi = (x\varphi)(y\varphi)$ for all $x, y \in S$ and $1_S\varphi = 1_T$.

Monomorphism, isomorphism

A *monomorphism* is an injective homomorphism. If $\varphi : S \to T$ is a surjective homomorphism, then T is a *homomorphic image* of S. An *isomorphism* is a bijective homomorphism. It is easy to prove that a homomorphism $\varphi : S \to T$ is an isomorphism if and only if there is a homomorphism $\varphi^{-1} : T \to S$ such that $\varphi\varphi^{-1} = \mathrm{id}_S$ and $\varphi^{-1}\varphi = \mathrm{id}_T$. If there is an isomorphism $\varphi : S \to T$, then we say S and T are *isomorphic* and denote this by $S \simeq T$.

When two semigroups are isomorphic, we can think of them as the 'same' abstract structure in different settings.

It is easy to prove that if $\varphi : S \to T$ is a homomorphism and S' and T' are subsemigroups of S and T respectively, then $S'\varphi$ is a subsemigroup of T and $T'\varphi^{-1}$ is a subsemigroup of S if it is non-empty. In particular, putting $S' = S$ shows that $\operatorname{im}\varphi$ is a subsemigroup of T. If φ is a monomorphism, then S is isomorphic to the subsemigroup $\operatorname{im}\varphi$ of T.

Kernel of a homomorphism

The *kernel* of a homomorphism $\varphi : S \to T$ is the binary relation

$$\ker\varphi = \{(x, y) \in S \times S : x\varphi = y\varphi\}.$$

Notice that φ is a monomorphism if and only if $\ker\varphi$ is the identity relation (that is, $\ker\varphi = \mathrm{id}_S$).

We now give a result showing that every semigroup is isomorphic to a subsemigroup of a semigroup of transformations. This is the analogue of Cayley's theorem for groups, which states that every group is isomorphic to a subgroup of a symmetric group. For any $x \in S$, let $\rho_x \in \mathcal{T}_{S^1}$ be the map defined by $s\rho_x = sx$ for all $s \in S^1$.

ρ_x

Right regular representation

THEOREM 1.22. *The map $\varphi : S \to \mathcal{T}_{S^1}$ given by $x \mapsto \rho_x$ is a monomorphism.*

Proof of 1.22. Let $x, y, s \in S$. Then $s\rho_x\rho_y = (sx)\rho_y = (sx)y = s(xy) = s\rho_{xy}$; hence $(x\varphi)(y\varphi) = \rho_x\rho_y = \rho_{xy} = (xy)\varphi$. Therefore φ is a homomorphism. Furthermore

$$x\varphi = y\varphi \Rightarrow \rho_x = \rho_y \Rightarrow 1\rho_x = 1\rho_y \Rightarrow 1x = 1y \Rightarrow x = y;$$

hence φ is injective. 1.22

End(S)

An *endomorphism* is a homomorphism from a semigroup to itself. The set of all endomorphisms of S is denoted $\mathrm{End}(S)$ and forms a subsemigroup of $\mathcal{T}_S$.

Group-embeddability

The semigroup S is *group-embeddable* if there exists a group G and a

monomorphism $\varphi : S \to G$. In this case, S is isomorphic to the subsemigroup $\operatorname{im}\varphi$ of G. Clearly any group-embeddable semigroup is cancellative, but we shall see that there exist cancellative semigroups that are not group-embeddable (see Example 2.14).

Anti-homomorphism

A map $\varphi : S \to T$ is an *anti-homomorphism* if $(xy)\varphi = (y\varphi)(x\varphi)$ for all $x, y \in S$.

CONGRUENCES AND QUOTIENTS

Congruence

A binary relation ρ on S is

- *left-compatible* if $(\forall x, y, z \in S)(x \rho y \Rightarrow zx \rho zy)$;
- *right-compatible* if $(\forall x, y, z \in S)(x \rho y \Rightarrow xz \rho yz)$;
- *compatible* if $(\forall x, y, z, t \in S)\big((x \rho y) \wedge (z \rho t) \Rightarrow xz \rho yt\big)$.

A left-compatible equivalence relation is a *left congruence*; a right-compatible equivalence relation is a *right congruence*; a compatible equivalence relation is a *congruence*.

Congruences are left/right congruences

PROPOSITION 1.23. *A relation ρ on S is a congruence if and only if it is both a left and a right congruence.*

Proof of 1.23. Suppose that the relation ρ is both a left and a right congruence. Let $x, y, z, t \in S$ be such that $x \rho y$ and $z \rho t$. Since ρ is a right congruence, $xz \rho yz$. Since ρ is a left congruence, $yz \rho yt$. Since ρ is transitive, $xz \rho yt$. Hence ρ is a congruence.

Suppose now that ρ is a congruence. Let $x, y \in S$ be such that $x \rho y$. Let $z \in S$. Since ρ is reflexive, $z \rho z$. Since ρ is a congruence, $zx \rho zy$ and $xz \rho yz$. Hence ρ is both a left and a right congruence. 1.23

Factor semigroup

Let ρ be a congruence on S. Let S/ρ denote the quotient set of S by ρ (that is, the set of ρ-classes of S). For any $x \in S$, let $[x]_\rho \in S/\rho$ be the ρ-class of x; that is, $[x]_\rho = \{ y \in S : y \rho x \}$. Define a multiplication on S/ρ by

$$[x]_\rho [y]_\rho = [xy]_\rho.$$

This multiplication is well-defined, in the sense that if we chose different representatives for the ρ-classes $[x]_\rho$ and $[y]_\rho$, we would get the same answer:

$$\begin{aligned} & ([x]_\rho = [x']_\rho) \wedge ([y]_\rho = [y']_\rho) \\ \Rightarrow\ & (x \rho x') \wedge (y \rho y') \\ \Rightarrow\ & xy \rho x'y' && \text{[since } \rho \text{ is a congruence]} \\ \Rightarrow\ & [xy]_\rho = [x'y']_\rho. \end{aligned}$$

The factor set S/ρ, with this multiplication, is a semigroup and is called the *quotient* or *factor* of S by ρ. The map $\rho^\natural : S \to S/\rho$, defined by $x\rho^\natural = [x]_\rho$ is clearly a surjective homomorphism, called the *natural map* or *natural homomorphism*.

Natural map

THEOREM 1.24. *Let $\varphi : S \to T$ be a homomorphism. Then $\ker\varphi$ is a congruence, and the map $\psi : S/\ker\varphi \to \operatorname{im}\varphi$ with $[x]_{\ker\varphi}\psi = x\varphi$ is an isomorphism, and so $S/\ker\varphi \simeq \operatorname{im}\varphi$.*

First isomorphism theorem

Proof of 1.24. Let $x, y, z, t \in S$. Then

$$\begin{aligned}
&(x, y) \in \ker\varphi \wedge (z, t) \in \ker\varphi \\
\Rightarrow\ &(x\varphi = y\varphi) \wedge (z\varphi = t\varphi) && \text{[by definition of } \ker\varphi\text{]} \\
\Rightarrow\ &(x\varphi)(z\varphi) = (y\varphi)(t\varphi) \\
\Rightarrow\ &(xz)\varphi = (yt)\varphi && \text{[since } \varphi \text{ is a homomorphism]} \\
\Rightarrow\ &(xz, yt) \in \ker\varphi; && \text{[by definition of } \ker\varphi\text{]}
\end{aligned}$$

thus $\ker\varphi$ is a congruence. Now,

$$\left.\begin{aligned}
&[x]_{\ker\varphi} = [y]_{\ker\varphi} \\
\Leftrightarrow\ &(x, y) \in \ker\varphi && \text{[by definition of } \ker\varphi\text{-classes]} \\
\Leftrightarrow\ &x\varphi = y\varphi && \text{[by definition of } \ker\varphi\text{]} \\
\Leftrightarrow\ &[x]_{\ker\varphi}\psi = [y]_{\ker\varphi}\psi. && \text{[by definition of } \psi\text{]}
\end{aligned}\right\} \quad (1.5)$$

The forward implication of (1.5) shows that ψ is well-defined. The reverse implication shows that ψ is injective. The image of ψ is clearly $\operatorname{im}\varphi$. The map ψ is a homomorphism since φ is a homomorphism. Hence ψ is an isomorphism and so $S/\ker\varphi \simeq \operatorname{im}\varphi$. 1.24

Let I be an ideal of S. Then $\rho_I = (I \times I) \cup \mathrm{id}_S$ is a congruence on S. The factor semigroup S/ρ_I is also denoted S/I, and the element $[x]_{\rho_I}$ is denoted $[x]_I$. The congruence ρ_I is called the *Rees congruence* induced by I, and S/I is a *Rees factor semigroup*. The elements of S/I are the ρ_I-classes, which comprise I and singleton sets $\{x\}$ for each $x \in S \setminus I$. It is easy to see that I is a zero of the factor semigroup S/I, so it is often convenient to view S/I as having elements $(S \setminus I) \cup \{0\}$, and to think of forming S/I by starting with S and merging all elements of I to form a zero; see Figure 1.5.

Rees factor semigroup

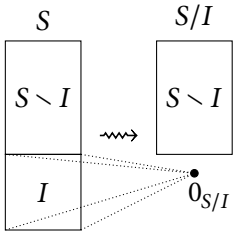


FIGURE 1.5
Forming S/I from S by merging elements of I to form a zero.

The following result shows that the ideals of S/I are in one-to-one correspondence with the ideals of S that contain I.

PROPOSITION 1.25. *Let I be an ideal of S. Let $\mathcal{A}$ be the collection of ideals of S that contain I. Let $\mathcal{B}$ be the collection of the ideals of S/I. Then the map $\varphi : \mathcal{A} \to \mathcal{B}$ given by $J\varphi = J/I$ is a bijection from $\mathcal{A}$ to $\mathcal{B}$ that preserves inclusion, in the sense that $J \subseteq J' \Rightarrow J\varphi \subseteq J'\varphi$.* 1.25

Ideal extension

A semigroup E is an *ideal extension* of S by T if S is an ideal of E and $E/S \simeq T$. Note that for an ideal extension of S by T to exist, T must contain a zero. Note further that there may be many non-isomorphic semigroups that are ideal extensions of S by T.

GENERATING EQUIVALENCES AND CONGRUENCES

In this section, we will study how an equivalence relation or congruence on a semigroup S is generated by a relation on S. This section is rather technical, but fundamentally important for future chapters.

Generating equivalences

Throughout this section, let X be a non-empty set. For any $\rho \in \mathcal{B}_X$, let

$$\begin{aligned}
\rho^{\mathsf{R}} &= \bigcap \{ \sigma \in \mathcal{B}_X : \rho \subseteq \sigma \wedge \sigma \text{ is reflexive} \};\\
\rho^{\mathsf{S}} &= \bigcap \{ \sigma \in \mathcal{B}_X : \rho \subseteq \sigma \wedge \sigma \text{ is symmetric} \};\\
\rho^{\mathsf{T}} &= \bigcap \{ \sigma \in \mathcal{B}_X : \rho \subseteq \sigma \wedge \sigma \text{ is transitive} \};\\
\rho^{\mathsf{E}} &= \bigcap \{ \sigma \in \mathcal{B}_X : \rho \subseteq \sigma \wedge \sigma \text{ is an equivalence relation} \}.
\end{aligned}$$

There is at least one element $\sigma \in \mathcal{B}_X$ fulfilling the condition in each of the collections above, namely $\sigma = X \times X$. Furthermore, since every element in these collections contains ρ, the intersections ρ^{R}, ρ^{S}, ρ^{T}, and ρ^{E} all contain ρ. It is easy to see that

- ρ^{R}, called the *reflexive closure of* ρ, is the smallest reflexive relation containing ρ;
- ρ^{S}, called the *symmetric closure of* ρ, is the smallest symmetric relation containing ρ;
- ρ^{T}, called the *transitive closure of* ρ, is the smallest transitive relation containing ρ;
- ρ^{E}, called the *equivalence relation generated by* ρ, is the smallest equivalence relation containing ρ.

PROPOSITION 1.26. *For any* $\rho \in \mathcal{B}_X$,

a) $\rho^{\mathsf{R}} = \rho \cup \mathrm{id}_X$;
b) $\rho^{\mathsf{S}} = \rho \cup \rho^{-1}$;
c) $\rho^{\mathsf{T}} = \bigcup_{n=1}^{\infty} \rho^n$;
d) $(\rho^{\mathsf{R}})^{\mathsf{S}} = (\rho^{\mathsf{S}})^{\mathsf{R}} = \rho \cup \rho^{-1} \cup \mathrm{id}_X$;
e) $(\rho^{\mathsf{R}})^{\mathsf{T}} = (\rho^{\mathsf{T}})^{\mathsf{R}} = \rho^{\mathsf{T}} \cup \mathrm{id}_X$;
f) $\rho^{\mathsf{E}} = ((\rho^{\mathsf{R}})^{\mathsf{S}})^{\mathsf{T}} = ((\rho^{\mathsf{S}})^{\mathsf{T}})^{\mathsf{R}} = \mathrm{id}_X \cup \bigcup_{n=1}^{\infty} (\rho \cup \rho^{-1})^n$.

Proof of 1.26. a) Since ρ^{R} is a reflexive relation containing ρ, it is immediate that $\rho \cup \mathrm{id}_X \subseteq \rho^{\mathsf{R}}$. On the other hand, $\rho \cup \mathrm{id}_X$ is a reflexive relation containing ρ; since ρ^{R} is the smallest reflexive relation containing ρ, we have $\rho^{\mathsf{R}} \subseteq \rho \cup \mathrm{id}_X$. Hence $\rho^{\mathsf{R}} = \rho \cup \mathrm{id}_X$.

b) Since ρ^{S} is a symmetric relation containing ρ, it is immediate that $\rho \cup \rho^{-1} \subseteq \rho^{\mathsf{S}}$. On the other hand, $\rho \cup \rho^{-1}$ is a symmetric relation containing ρ; since ρ^{S} is the smallest symmetric relation containing ρ, we have $\rho^{\mathsf{S}} \subseteq \rho \cup \rho^{-1}$. Hence $\rho^{\mathsf{S}} = \rho \cup \rho^{-1}$.

c) Since ρ^{T} contains ρ, transitivity implies that it contains $\rho^2 = \rho \circ \rho$. Transitivity again implies that ρ^{T} contains $\rho^3 = \rho \circ \rho^2$. Continuing inductively, we see that ρ^{T} contains ρ^n for all $n \in \mathbb{N}$; hence $\bigcup_{n=1}^{\infty} \rho^n \subseteq \rho^{\mathsf{T}}$.

On the other hand,

$$\begin{aligned}
&(x, y), (y, z) \in \textstyle\bigcup_{n=1}^{\infty} \rho^n \\
\Rightarrow\ &(\exists k, \ell \in \mathbb{N})\big((x, y) \in \rho^k \wedge (y, z) \in \rho^\ell\big) \\
\Rightarrow\ &(\exists k, \ell \in \mathbb{N})\big((x, z) \in \rho^k \circ \rho^\ell\big) \\
\Rightarrow\ &(\exists k, \ell \in \mathbb{N})\big((x, z) \in \rho^{k+\ell}\big) \\
\Rightarrow\ &(x, z) \in \textstyle\bigcup_{n=1}^{\infty} \rho^n.
\end{aligned}$$

So $\bigcup_{n=1}^{\infty} \rho^n$ is a transitive relation containing ρ. Since ρ^{T} is the smallest such relation, we have $\rho^{\mathsf{T}} \subseteq \bigcup_{n=1}^{\infty} \rho^n$. Hence $\rho^{\mathsf{T}} = \bigcup_{n=1}^{\infty} \rho^n$.

d) We have

$$\begin{aligned}
&(\rho^{\mathsf{R}})^{\mathsf{S}} && && (1.6)\\
&= \rho^{\mathsf{R}} \cup (\rho^{\mathsf{R}})^{-1} && \text{[by part b)]} \\
&= \rho \cup \mathrm{id}_X \cup (\rho \cup \mathrm{id}_X)^{-1} && \text{[by part a)]} \\
&= \rho \cup \mathrm{id}_X \cup \rho^{-1} \cup \mathrm{id}_X^{-1} && \text{[by definition of converse]} \\
&= \rho \cup \rho^{-1} \cup \mathrm{id}_X && [\text{since } \mathrm{id}_X^{-1} = \mathrm{id}_X] && (1.7)\\
&= \rho^{\mathsf{S}} \cup \mathrm{id}_X && \text{[by part b)]} \\
&= (\rho^{\mathsf{S}})^{\mathsf{R}}. && \text{[by part a)]} && (1.8)
\end{aligned}$$

The result is given by lines (1.6), (1.7), and (1.8).

e) Since $\mathrm{id}_X \subseteq \rho^{\mathsf{R}}$, we have $\mathrm{id}_X^{\mathsf{T}} \subseteq (\rho^{\mathsf{R}})^{\mathsf{T}}$. Since $\mathrm{id}_X \circ \mathrm{id}_X = \mathrm{id}_X$, it follows from part c) that $\mathrm{id}_X^{\mathsf{T}} = \mathrm{id}_X$. So $\mathrm{id}_X \subseteq (\rho^{\mathsf{R}})^{\mathsf{T}}$. Since $\rho \subseteq \rho^{\mathsf{R}}$, we have $\rho^{\mathsf{T}} \subseteq (\rho^{\mathsf{R}})^{\mathsf{T}}$. So $(\rho^{\mathsf{T}})^{\mathsf{R}} = \rho^{\mathsf{T}} \cup \mathrm{id}_X \subseteq (\rho^{\mathsf{R}})^{\mathsf{T}}$.

Now let $(u, v) \in (\rho^{\mathsf{R}})^{\mathsf{T}}$. Then by parts a) and c), $(u, v) \in \bigcup_{n=1}^{\infty} (\rho \cup \mathrm{id}_X)^n$. So there exists $n \in \mathbb{N}$ and $x_0, \ldots, x_n \in X$ such that $u = x_0$, $v = x_n$, and $(x_i, x_{i+1}) \in \rho \cup \mathrm{id}_X$ for $i = 0, \ldots, n-1$. Fix such a sequence with n minimal. Then if $n \geqslant 2$, no pair (x_i, x_{i+1}) is in id_X, for this would imply that $x_i = x_{i+1}$ and so we could shorten the sequence by

deleting one of x_i or x_{i+1}, contradicting the minimality of n. So

$$(u,v) \in \mathrm{id}_X \cup \rho \cup \bigcup_{n=2}^{\infty} \rho^n = \mathrm{id}_X \cup \rho^{\mathsf{T}} = (\rho^{\mathsf{T}})^{\mathsf{R}}.$$

Hence $(\rho^{\mathsf{R}})^{\mathsf{T}} \subseteq (\rho^{\mathsf{T}})^{\mathsf{R}}$ and so $(\rho^{\mathsf{R}})^{\mathsf{T}} = (\rho^{\mathsf{T}})^{\mathsf{R}}$.

f) Since ρ^{E} is reflexive and contains ρ, it contains ρ^{R}. Since it is symmetric and contains ρ^{R}, it contains $(\rho^{\mathsf{R}})^{\mathsf{S}}$. Since it is transitive and contains $(\rho^{\mathsf{R}})^{\mathsf{S}}$, it contains $((\rho^{\mathsf{R}})^{\mathsf{S}})^{\mathsf{T}}$. Hence $((\rho^{\mathsf{R}})^{\mathsf{S}})^{\mathsf{T}} \subseteq \rho^{\mathsf{E}}$.

On the other hand, $((\rho^{\mathsf{R}})^{\mathsf{S}})^{\mathsf{T}}$ is transitive by the definition of $^{\mathsf{T}}$. Furthermore, $((\rho^{\mathsf{R}})^{\mathsf{S}})^{\mathsf{T}} \supseteq (\rho^{\mathsf{R}})^{\mathsf{S}}) \supseteq \rho^{\mathsf{R}} \supseteq \mathrm{id}_X$ and so $((\rho^{\mathsf{R}})^{\mathsf{S}})^{\mathsf{T}}$ is reflexive. Let $(x,y) \in ((\rho^{\mathsf{R}})^{\mathsf{S}})^{\mathsf{T}} = \bigcup_{n=1}^{\infty}((\rho^{\mathsf{R}})^{\mathsf{S}})^n$. Then $(x,y) \in ((\rho^{\mathsf{R}})^{\mathsf{S}})^n$ for some $n \in \mathbb{N}$. Hence there exist $x_0, \ldots, x_n \in X$ with $x_0 = x$, $x_n = y$, and $(x_i, x_{i+1}) \in (\rho^{\mathsf{R}})^{\mathsf{S}}$ for $i = 0, \ldots, n-1$. Since $(\rho^{\mathsf{R}})^{\mathsf{S}}$ is symmetric, $(x_{i+1}, x_i) \in (\rho^{\mathsf{R}})^{\mathsf{S}}$ for each i, and so $(y,x) \in ((\rho^{\mathsf{R}})^{\mathsf{S}})^n \subseteq ((\rho^{\mathsf{R}})^{\mathsf{S}})^{\mathsf{T}}$. So $((\rho^{\mathsf{R}})^{\mathsf{S}})^{\mathsf{T}}$ is symmetric. Hence $((\rho^{\mathsf{R}})^{\mathsf{S}})^{\mathsf{T}}$ is an equivalence relation containing ρ. Since ρ^{E} is the smallest equivalence relation containing ρ, we have $\rho^{\mathsf{E}} \subseteq ((\rho^{\mathsf{R}})^{\mathsf{S}})^{\mathsf{T}}$. Hence $\rho^{\mathsf{E}} = ((\rho^{\mathsf{R}})^{\mathsf{S}})^{\mathsf{T}}$.

Finally, notice that

$$\begin{aligned}
\rho^{\mathsf{E}} &= ((\rho^{\mathsf{R}})^{\mathsf{S}})^{\mathsf{T}} && \text{[by the above reasoning]} && (1.9)\\
&= ((\rho^{\mathsf{S}})^{\mathsf{R}})^{\mathsf{T}} && \text{[by part d)]}\\
&= ((\rho^{\mathsf{S}})^{\mathsf{T}})^{\mathsf{R}} && \text{[by part e)]} && (1.10)\\
&= \mathrm{id}_X \cup (\rho^{\mathsf{S}})^{\mathsf{T}} && \text{[by part a)]}\\
&= \mathrm{id}_X \cup (\rho \cup \rho^{-1})^{\mathsf{T}} && \text{[by part b)]}\\
&= \mathrm{id}_X \cup \textstyle\bigcup_{n=1}^{\infty}(\rho \cup \rho^{-1}). && \text{[by part c)]} && (1.11)
\end{aligned}$$

Lines (1.9), (1.10), and (1.11) give the three required equalities. 1.26

Generating congruences

For any $\rho \in \mathcal{B}_S$, let

$$\rho^{\mathsf{C}} = \bigcap\{\sigma \in \mathcal{B}_S : \rho \subseteq \sigma \wedge \sigma \text{ is left and right compatible}\},$$
$$\rho^{\#} = \bigcap\{\sigma \in \mathcal{B}_S : \rho \subseteq \sigma \wedge \sigma \text{ is a congruence}\}.$$

It is easy to see that

- ρ^{C} is the smallest left and right compatible relation containing ρ;
- $\rho^{\#}$, called the *congruence generated by* ρ, is the smallest congruence containing ρ.

PROPOSITION 1.27. *For any $\rho \in \mathcal{B}_S$, we have $\rho^{\mathsf{C}} = \{(pxq, pyq) \in S \times S : p, q \in S^1 \wedge (x,y) \in \rho\}$.*

Proof of 1.27. Let $\sigma = \{(pxq, pyq) \in S \times S : p, q \in S^1, (x,y) \in \rho\}$. To prove that $\sigma = \rho^{\mathsf{C}}$, we have to show that σ is the smallest left and right compatible relation on S containing ρ. Notice first that if $(x,y) \in \rho$, then

$(x, y) = (1x1, 1y1) \in \sigma$. Hence σ contains ρ. Let $(u, v) \in \sigma$ and $r \in S$. Then $u = pxq$ and $v = pyq$ for some $(x, y) \in \rho$. Let $p' = rp$. Then $(ru, rv) = (p'xq, p'yq) \in \sigma$. Hence σ is left-compatible. Similarly, σ is right compatible.

Now let τ be some left and right compatible relation that contains ρ. Let $(pxq, pyq) \in \sigma$, where $(x, y) \in \rho$ and $p, q \in S^1$. Then $(x, y) \in \tau$ since $\rho \subseteq \tau$. Hence $(pxq, pyq) \in \tau$ since τ is left and right compatible. Thus $\sigma \subseteq \tau$. Therefore σ is the smallest left and right compatible relation containing ρ. 1.27

PROPOSITION 1.28. *For any* $\rho, \sigma \in \mathcal{B}_S$,

a) $(\rho \cup \sigma)^{\mathsf{C}} = \rho^{\mathsf{C}} \cup \sigma^{\mathsf{C}}$;

b) $(\rho^{-1})^{\mathsf{C}} = (\rho^{\mathsf{C}})^{-1}$.

Proof of 1.28. a) For $u, v \in S$,

$$\begin{aligned}
&(u, v) \in (\rho \cup \sigma)^{\mathsf{C}} \\
\Leftrightarrow\ &(\exists p, q \in S^1, (x, y) \in \rho \cup \sigma)(u = pxq \wedge v = pyq) \\
&\qquad\qquad \text{[by Proposition 1.27]} \\
\Leftrightarrow\ &(\exists p, q \in S^1, (x, y) \in \rho)(u = pxq \wedge v = pyq) \\
&\qquad \vee (\exists p, q \in S^1, (x, y) \in \sigma)(u = pxq \wedge v = pyq) \\
\Leftrightarrow\ &(u, v) \in \rho^{\mathsf{C}} \vee (u, v) \in \sigma^{\mathsf{C}} \qquad \text{[by Proposition 1.27]} \\
\Leftrightarrow\ &(u, v) \in \rho^{\mathsf{C}} \cup \sigma^{\mathsf{C}}.
\end{aligned}$$

b) For $u, v \in S$,

$$\begin{aligned}
&(u, v) \in (\rho^{-1})^{\mathsf{C}} \\
\Leftrightarrow\ &(\exists p, q \in S^1, (x, y) \in \rho^{-1})(u = pxq \wedge v = pyq) \\
&\qquad\qquad \text{[by Proposition 1.27]} \\
\Leftrightarrow\ &(\exists p, q \in S^1, (y, x) \in \rho)(v = pyq \wedge u = pxq) \\
\Leftrightarrow\ &(v, u) \in \rho^{\mathsf{C}} \qquad \text{[by Proposition 1.27]} \\
\Leftrightarrow\ &(u, v) \in (\rho^{\mathsf{C}})^{-1}.
\end{aligned}$$

1.28

PROPOSITION 1.29. *For any* $\rho \in \mathcal{B}_S$,

Characterizing generated congruences

$$\rho^{\#} = (\rho^{\mathsf{C}})^{\mathsf{E}} = \mathrm{id}_S \cup \bigcup_{n=1}^{\infty} (\rho^{\mathsf{C}} \cup (\rho^{\mathsf{C}})^{-1})^n.$$

Proof of 1.29. By Proposition 1.26(f),

$$(\rho^{\mathsf{C}})^{\mathsf{E}} = \mathrm{id}_S \cup \bigcup_{n=1}^{\infty} (\rho^{\mathsf{C}} \cup (\rho^{\mathsf{C}})^{-1})^n,$$

so we must prove that $\rho^{\#} = (\rho^{\mathsf{C}})^{\mathsf{E}}$. That is, we must show that $(\rho^{\mathsf{C}})^{\mathsf{E}}$ is the smallest congruence containing ρ. By definition, $(\rho^{\mathsf{C}})^{\mathsf{E}}$ is an equivalence relation containing ρ^{C}, which in turn contains ρ. So $\rho \subseteq (\rho^{\mathsf{C}})^{\mathsf{E}}$.

Now let $x, y, z \in S$ and suppose that $(x, y) \in (\rho^{\mathsf{C}})^{\mathsf{E}}$. If $(x, y) \in \mathrm{id}_S$, then $x = y$, and so $zx = zy$, and thus $(zx, zy) \in \mathrm{id}_S \subseteq (\rho^{\mathsf{C}})^{\mathsf{E}}$. Furthermore,

$$\begin{aligned}
&(x, y) \in \textstyle\bigcup_{n=1}^{\infty} (\rho^{\mathsf{C}} \cup (\rho^{\mathsf{C}})^{-1})^n \\
\Rightarrow\ &(x, y) \in \textstyle\bigcup_{n=1}^{\infty} ((\rho \cup \rho^{-1})^{\mathsf{C}})^n && \text{[by Proposition 1.28]} \\
\Rightarrow\ &(\exists n \in \mathbb{N})(\exists x_0, x_1, \ldots, x_n \in S) \\
&\qquad [(x = x_0) \wedge (x_n = y) \\
&\qquad\qquad \wedge (\forall i)((x_i, x_{i+1}) \in (\rho \cup \rho^{-1})^{\mathsf{C}})] && \text{[by definition of } \circ] \\
\Rightarrow\ &(\exists n \in \mathbb{N})(\exists x_0, x_1, \ldots, x_n \in S) \\
&\qquad [(zx = zx_0) \wedge (zx_n = zy) \\
&\qquad\qquad \wedge (\forall i)((zx_i, zx_{i+1}) \in (\rho \cup \rho^{-1})^{\mathsf{C}})] \\
&&& \text{[since } (\rho \cup \rho^{-1})^{\mathsf{C}} \text{ is left and right compatible]} \\
\Rightarrow\ &(\exists n \in \mathbb{N})((zx, zy) \in ((\rho \cup \rho^{-1})^{\mathsf{C}})^n) && \text{[by definition of } \circ] \\
\Rightarrow\ &(zx, zy) \in \textstyle\bigcup_{n=1}^{\infty} (\rho^{\mathsf{C}} \cup (\rho^{\mathsf{C}})^{-1})^n && \text{[by Proposition 1.28]} \\
\Rightarrow\ &(zx, zy) \in (\rho^{\mathsf{C}})^{\mathsf{E}}.
\end{aligned}$$

Hence $(x, y) \in (\rho^{\mathsf{C}})^{\mathsf{E}}$ implies $(zx, zy) \in (\rho^{\mathsf{C}})^{\mathsf{E}}$. Therefore $(\rho^{\mathsf{C}})^{\mathsf{E}}$ is left-compatible. Similarly, $(\rho^{\mathsf{C}})^{\mathsf{E}}$ is right-compatible. Hence $(\rho^{\mathsf{C}})^{\mathsf{E}}$ is a congruence containing ρ.

Now suppose that τ is a congruence containing ρ. Then τ is left and right compatible and so must contain ρ^{C}, which is the smallest left and right compatible relation containing ρ. Furthermore, τ is an equivalence relation, and so it must contain $(\rho^{\mathsf{C}})^{\mathsf{E}}$, which is the smallest equivalence relation containing ρ^{C}. Hence $(\rho^{\mathsf{C}})^{\mathsf{E}} \subseteq \tau$. Therefore $(\rho^{\mathsf{C}})^{\mathsf{E}}$ is the smallest congruence containing ρ. 1.29

Lattice of congruences

Let $\mathcal{E}_S$ be the set of equivalence relations on S and let $\mathcal{C}_S$ be the set of congruences on S. Then $\mathcal{E}_S$ and $\mathcal{C}_S$ both admit $\subseteq$ as a partial order. It is easy to see that both $(\mathcal{E}_S, \subseteq)$ and $(\mathcal{C}_S, \subseteq)$ are actually lattices:

- for any $\rho, \sigma \in \mathcal{E}_S$, we have $\rho \sqcap \sigma = \rho \cap \sigma$ and $\rho \sqcup \sigma = (\rho \cup \sigma)^{\mathsf{E}}$;
- for any $\rho, \sigma \in \mathcal{C}_S$, we have $\rho \sqcap \sigma = \rho \cap \sigma$ and $\rho \sqcup \sigma = (\rho \cup \sigma)^{\#}$.

Suppose $\rho, \sigma \in \mathcal{C}_S$. There seems to be an ambiguity in writing $\rho \sqcup \sigma$: do we mean the join $(\rho \cup \sigma)^{\mathsf{E}}$ in the lattice of equivalence relations $\mathcal{E}_S$, or the join $(\rho \cup \sigma)^{\#}$ in the lattice of congruences $\mathcal{C}_S$? However,

$$\begin{aligned}
&(\rho \cup \sigma)^{\#} \\
&= ((\rho \cup \sigma)^{\mathsf{C}})^{\mathsf{E}} && \text{[by Proposition 1.29]} \\
&= (\rho^{\mathsf{C}} \cup \sigma^{\mathsf{C}})^{\mathsf{E}} && \text{[by Proposition 1.28(a)]} \\
&= (\rho \cup \sigma)^{\mathsf{E}}. && \text{[since } \rho \text{ and } \sigma \text{ are compatible]}
\end{aligned}$$

So there is really no ambiguity in writing $\rho \sqcup \sigma$.

PROPOSITION 1.30. *Let $\rho, \sigma \in \mathcal{E}_S$. Then $\rho \sqcup \sigma = (\rho \circ \sigma)^{\mathsf{T}}$.*

Characterizing join of equivalence relations

Proof of 1.30. Since $\rho \cup \sigma$ contains both ρ and σ, it follows that

$$\rho \circ \sigma \subseteq (\rho \cup \sigma) \circ (\rho \cup \sigma) = (\rho \cup \sigma)^2,$$

and more generally that $(\rho \circ \sigma)^n \subseteq (\rho \cup \sigma)^{2n}$. Thus

$$(\rho \circ \sigma)^{\mathsf{T}} = \bigcup_{n=1}^{\infty} (\rho \circ \sigma)^n \subseteq \bigcup_{n=1}^{\infty} (\rho \cup \sigma)^n = (\rho \cup \sigma)^{\mathsf{T}}. \tag{1.12}$$

On the other hand, $\rho \circ \sigma$ contains $\rho \circ \mathrm{id}_S = \rho$ (since σ is reflexive) and contains $\mathrm{id}_S \circ \sigma = \sigma$ (since ρ is reflexive), and thus $\rho \cup \sigma \subseteq \rho \circ \sigma$. Hence $(\rho \cup \sigma)^{\mathsf{T}} \subseteq (\rho \circ \sigma)^{\mathsf{T}}$. Combine this with (1.12) to see that

$$(\rho \cup \sigma)^{\mathsf{T}} = (\rho \circ \sigma)^{\mathsf{T}}. \tag{1.13}$$

Then

$$\begin{aligned}
&\rho \sqcup \sigma \\
&= (\rho \cup \sigma)^{\mathsf{E}} \\
&= (((\rho \cup \sigma)^{\mathsf{R}})^{\mathsf{S}})^{\mathsf{T}} && \text{[by Proposition 1.26(f)]} \\
&= ((\rho \cup \sigma) \cup (\rho \cup \sigma)^{-1} \cup \mathrm{id}_S)^{\mathsf{T}} && \text{[by Proposition 1.26(d)]} \\
&= (\rho \cup \sigma \cup \rho^{-1} \cup \sigma^{-1} \cup \mathrm{id}_S)^{\mathsf{T}} \\
&= (\rho \cup \sigma)^{\mathsf{T}} && \text{[since } \rho \text{ and } \sigma \text{ are reflexive and symmetric]} \\
&= (\rho \circ \sigma)^{\mathsf{T}}. && \text{[by (1.13)]}
\end{aligned}$$

This completes the proof. 1.30

PROPOSITION 1.31. *Let $\rho, \sigma \in \mathcal{E}_S$. If $\rho \circ \sigma = \sigma \circ \rho$, then $\rho \sqcup \sigma = \rho \circ \sigma$.*

Join of commuting equivalence relations

Proof of 1.31. Suppose $\rho \circ \sigma = \sigma \circ \rho$. Then

$$(\rho \circ \sigma)^2 = \rho \circ \sigma \circ \rho \circ \sigma = \rho \circ \rho \circ \sigma \circ \sigma = \rho^2 \circ \sigma^2. \tag{1.14}$$

But $\rho^2 \subseteq \rho$ and $\sigma^2 \subseteq \sigma$ since ρ and σ are transitive. Furthermore, $\rho = \rho \circ \mathrm{id}_S \subseteq \rho^2$ since ρ is reflexive, and similarly $\sigma \subseteq \sigma^2$. Hence $\rho^2 = \rho$ and $\sigma^2 = \sigma$ and so $(\rho \circ \sigma)^2 = \rho \circ \sigma$ by (1.14). Hence $(\rho \circ \sigma)^n = \rho \circ \sigma$ for all $n \in \mathbb{N}$, and thus

$$\begin{aligned}
\rho \sqcup \sigma &= (\rho \circ \sigma)^{\mathsf{T}} && \text{[by Proposition 1.30]} \\
&= \textstyle\bigcup_{n=1}^{\infty} (\rho \circ \sigma)^n && \text{[by Proposition 1.26(c)]} \\
&= \rho \circ \sigma. && \boxed{1.31}
\end{aligned}$$

SUBDIRECT PRODUCTS

Let $\mathcal{S} = \{ S_i : i \in I \}$ be a collection of semigroups. For each $j \in I$, there is a projection map from the direct product $\prod_{i \in I} S_i$ to S_j, taking an element of the direct product to its j-th component:

$$\pi_j : \prod_{i \in I} S_i \to S_j, \qquad x\pi_j = (j)x.$$

Notice that every π_j is a surjective homomorphism.

Subdirect product

A *subdirect product* of $\mathcal{S}$ is [a semigroup isomorphic to] a subsemigroup P of the direct product $\prod_{i \in I} S_i$ such that $P\pi_j = S_j$ for all $j \in I$.

Separation by surjective homomorphisms

Let S be a semigroup. A collection of surjective homomorphisms $\Phi = \{ \varphi_i : S \to S_i : i \in I \}$ is said to *separate* the elements of S if they have the property that

$$(\forall i \in I)(x\varphi_i = y\varphi_i) \Rightarrow x = y.$$

PROPOSITION 1.32. *A semigroup S is a subdirect product of a collection of semigroups $\mathcal{S} = \{ S_i : i \in I \}$ if and only if there exists a collection of surjective homomorphisms $\Phi = \{ \varphi_i : S \to S_i : i \in I \}$ that separate the elements of S.*

Proof of 1.32. If S is a subdirect product of $\mathcal{S}$, then the collection of projection maps restricted to S (that is, the collection $\{ \pi_i|_S : S \to S_i : i \in I \}$) separates the elements of S.

On the other hand, suppose the collection Φ separates the elements of S. Define $\psi : S \to \prod_{i \in I} S_i$ by letting the i-th component of $s\psi$ be $s\varphi_i$; that is, $(i)(s\psi) = s\varphi_i$. Then ψ is a homomorphism since each φ_i is a homomorphism. Furthermore, $s\psi = t\psi$ implies that $s\varphi_i = t\varphi_i$ for all $i \in I$, which implies $s = t$ since Φ separates the elements of S. Hence ψ is injective. So S is isomorphic to the subsemigroup $\operatorname{im} \psi$ of $\prod_{i \in I} S_i$. Finally, the projection maps π_i are all surjective since each φ_i is surjective. So $\operatorname{im} \varphi$ is a subdirect product of $\mathcal{S}$. 1.32

PROPOSITION 1.33. *Let S be a semigroup and let $\Sigma = \{ \sigma_i : i \in I \}$ be a collection of congruences on S. Let $\sigma = \bigcap \Sigma$. Then S/σ is a subdirect product of $\{ S/\sigma_i : i \in I \}$.*

Proof of 1.33. For each i there is a homomorphism $\varphi_i : S/\sigma \to S/\sigma_i$ with $[x]_\sigma \varphi_i = [x]_{\sigma_i}$. (These maps are well-defined since $\sigma \subseteq \sigma_i$.) Clearly, the homomorphisms φ_i are surjective. Furthermore, the collection $\Phi = \{ \varphi_i : i \in I \}$ separates the elements of S/σ, since if $[x]_\sigma \varphi_i = [y]_\sigma \varphi_i$ for all $i \in I$, then $[x]_{\sigma_i} = [y]_{\sigma_i}$ and thus $(x, y) \in \sigma_i$ for all $i \in I$, which implies $(x, y) \in \sigma = \bigcap \Sigma$ and so $[x]_\sigma = [y]_\sigma$. Therefore S/σ is a subdirect product of $\{ S/\sigma_i : i \in I \}$ by Proposition 1.32. 1.33

Ideal extensions of monoids are subdirect products

PROPOSITION 1.34. *Let M be a monoid and let E be an ideal extension of M by a semigroup T. Then E is a subdirect product of M and T.*

Proof of 1.34. By definition, M is an ideal of E and T is the Rees factor semigroup E/M. Let $\varphi : E \to T$ be the natural homomorphism $x\varphi = [x]_M$. Let $\psi : E \to M$ be given by $x\psi = x1_M$. Then

$$\begin{aligned}(x\psi)(y\psi) &= x1_M y1_M \\ &= xy1_M && \text{[since } y1_M \text{ lies in the ideal } M \text{ of } E\text{]} \\ &= (xy)\psi.\end{aligned}$$

Thus ψ is a homomorphism. Both φ and ψ are clearly surjective. Furthermore, if $x\varphi = y\varphi$ and $x\psi = y\psi$, then either $x, y \in E \setminus M$ and $[x]_M = [y]_M$ and so $x = y$, or $x, y \in M$ and $x1_M = y1_M$ and so $x = y$. Thus the collection of surjective homomorphisms $\{\varphi, \psi\}$ separates elements of E and so E is a subdirect product of M and T. 1.34

ACTIONS

Semigroup action

A *semigroup action* of a semigroup S on a set A is an operation $\cdot : A \times S \to A$ that is compatible with the semigroup multiplication, in the sense that

$$(a \cdot x) \cdot y = a \cdot (xy) \tag{1.15}$$

for all $a \in A$ and $x, y \in S$. We call such a semigroup action an *action of S on A*, or an *S-action on A*, and say that *S acts on A*.

EXAMPLE 1.35. a) Any subsemigroup S of $\mathcal{T}_A$ acts on A by $a \cdot \rho = a\rho$ (where $\rho \in \mathcal{T}_A$).

b) Let S be a subsemigroup of a semigroup T. Then S acts on T via $t \cdot x = tx$ for all $t \in T$ and $x \in S$. In particular, this holds when $T = S$ or when $T = S^1$.

Given an action $\cdot$, we can define a map $\varphi : S \to \mathcal{T}_A$, where the transformation $s\varphi$ is such that $a(s\varphi) = a \cdot s$. The condition (1.15) implies that φ is a homomorphism. Conversely, given a homomorphism $\varphi : S \to \mathcal{T}_A$, we can define an action $\cdot$ by $a \cdot s = a(s\varphi)$, which satisfies (1.15) since φ is a homomorphism. There is thus a one-to-one correspondence between actions of a semigroup S on A and homomorphisms $\varphi : S \to \mathcal{T}_A$.

Free, transitive, regular actions

An action of S on A is *free* if distinct elements of S act differently on every element of A, or, equivalently,

$$(\forall x, y \in S)\big((\exists a \in A)(a \cdot x = a \cdot y) \Rightarrow x = y\big).$$

An action of S on A is *transitive* if A is non-empty and for all $a, b \in A$, there is some element $s \in S$ such that $a \cdot s = b$. That is, the action is transitive if one can reach start at any element of A and reach any element

(possibly the same one) by acting by some element of S. An action is *regular* if it is both free and transitive. It is easy to see that if S has a regular action on A, then $|S| = |A|$.

Action by endomorphisms

Suppose A is also a semigroup. An action of S on A is *by endomorphisms* if $s\varphi \in \operatorname{End} A$ for each $s \in S$; in this case,

$$ab \cdot x = (a \cdot x)(b \cdot x)$$

for all $a, b \in A$ and $x \in S$.

Left action

The above discussions concern *right* semigroup actions. There is a dual notion of a *left semigroup action* of S on A, which is an operation $\cdot : S \times A \to A$ satisfying

$$s \cdot (t \cdot a) = (st) \cdot a;$$

this corresponds to a map $\varphi : S \to \mathcal{T}_A$, where $a(s\varphi) = s \cdot a$. This map φ is an anti-homomorphism since

$$a(t\varphi)(s\varphi) = (t \cdot a)(s\varphi) = s \cdot (t \cdot a) = st \cdot a = a((st)\varphi).$$

The definitions of actions being free, transitive, regular, and by endomorphisms also apply to left actions.



The correspondence of right actions with homomorphisms and left actions with anti-homomorphisms depends on writing maps on the right and composing them left-to-right. When maps are written on the left and composed right-to-left, right actions correspond to anti-homomorphisms and left actions to homomorphisms.

CAYLEY GRAPHS

Let S be a semigroup or monoid with a generating set A. The *right* (respectively, *left*) *Cayley graph* $\Gamma(S, A)$ (respectively, $\Gamma'(S, A)$) of S with respect to A is the directed graph with vertex set S and, for every $x \in S$ and $a \in A$, an edge from x to xa (respectively, ax) labelled by a. By default 'Cayley graph' means 'right Cayley graph'.

EXAMPLE 1.36. a) Let M be the monoid $(\mathbb{N} \cup \{0\}) \times (\mathbb{N} \cup \{0\})$. Let $a = (1, 0)$ and $b = (0, 1)$ and let $A = \{a, b\}$. The Cayley graph $\Gamma(M, A)$ is an infinite grid; part of it is shown in Figure 1.6.

b) Let $X = \{1, 2\}$. Let $A = \{\sigma, \pi, \upsilon\} \subseteq \mathcal{P}_X$, where

$$\sigma = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}, \quad \pi = \begin{pmatrix} 1 & 2 \\ 1 & 1 \end{pmatrix}, \quad \text{and } \upsilon = \begin{pmatrix} 1 & 2 \\ 1 & * \end{pmatrix}.$$

Then A generates $\mathcal{P}_X$. The Cayley graph $\Gamma(\mathcal{P}_X, A)$ is shown in Figure 1.7. Note the subgroup $\mathcal{S}_X$ and the subsemigroup $\mathcal{T}_X$, and that $\begin{pmatrix} 1 & 2 \\ * & * \end{pmatrix}$ is a zero and a sink vertex of the graph.

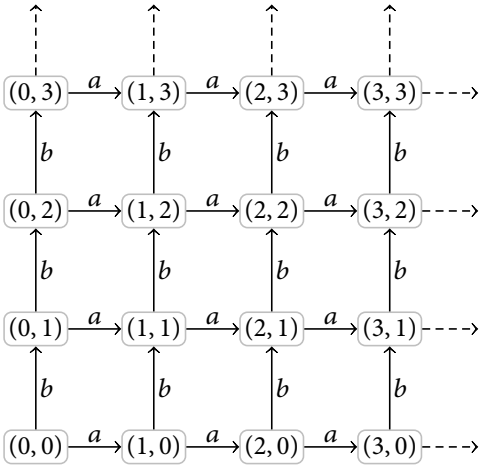


FIGURE 1.6
Cayley graph of $M = (\mathbb{N} \cup \{0\}) \times (\mathbb{N} \cup \{0\})$.

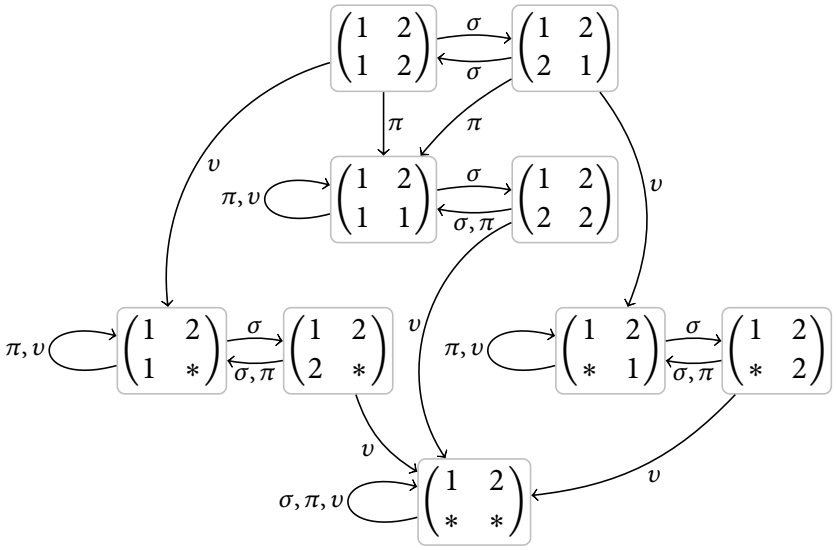


FIGURE 1.7
Cayley graph of $\mathcal{P}_{\{1,2\}}$.

c) Let $S = \{x, y\}$ be a two-element right zero semigroup and let $A = S$. The i) right and ii) left Cayley graphs $\Gamma(S, A)$ and $\Gamma'(S, A)$ are shown in Figure 1.8.

For groups, Cayley graphs have special properties. First, the left and right Cayley graphs are isomorphic under the map sending each vertex and each edge label to its inverse. Second, the Cayley graphs are connected, and indeed strongly connected. Third, the Cayley graphs are homogeneous, which essentially means that a neighbourhood of any vertex 'looks like' the corresponding neighbourhood of any other vertex. The graphs in Example 1.36(c) show that the left and right Cayley graphs of a semigroup need not be isomorphic; the second graph shows that the Cayley graph of a semigroup need not be connected. All the graphs in Example 1.36 except (c)(ii) show that Cayley graphs of semigroups need not be homogeneous.

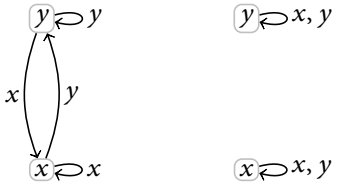


FIGURE 1.8
Right (i) and left (ii) Cayley graphs of a two-element right zero semigroup $\{x, y\}$.

EXERCISES

[See pages 203–209 for the solutions.]

1.1 Prove that if S is a semigroup and $e \in S$ is both a right zero and a right identity, then S is trivial.

1.2 Prove the following:

a) If S is a monoid with identity 1, the semigroup S^0 obtained by adjoining a zero if necessary is also a monoid with identity 1.

b) If S is a semigroup with zero 0, the monoid S^1 obtained by adjoining an identity if necessary also has zero 0.

✶1.3 Let S be a left-cancellative semigroup. Suppose that $e \in S$ is an idempotent. Prove that e is a left identity. Deduce that a cancellative semigroup can contain at most one idempotent, which must be an identity.

✶1.4 Prove that a right zero semigroup is left-cancellative.

✶1.5 Prove that a finite cancellative semigroup is a group.

1.6 Prove from the definition that id_X is an identity for $\mathcal{B}_X$. Does $\mathcal{B}_X$ contain a zero?

1.7 Does there exist a non-trivial semigroup that does not contain any proper subsemigroups?

✶1.8 Give an [easy] example of an infinite periodic semigroup.

1.9 Does either $\mathcal{T}_X$ or $\mathcal{P}_X$ contain a zero? A left zero? A right zero? [Note that the answer may depend on $|X|$.]

Power semigroup

1.10 The *power semigroup* of a semigroup S is the set $\mathbb{P}S$ of all subsets of S under the operation $XY = \{xy : x \in X, y \in Y\}$ for $X, Y \in \mathbb{P}S$. (Recall from page 5 that $X(YZ) = (XY)Z$ for all $X, Y, Z \in \mathbb{P}S$.)

a) Prove that $\mathbb{P}S$ contains a subsemigroup isomorphic to S.

b) Prove that $\varnothing$ is a zero of $\mathbb{P}S$. Prove that $(\mathbb{P}S) \setminus \{\varnothing\}$ is a subsemigroup of $\mathbb{P}S$.

c) Let M be a monoid. Prove that $(\mathbb{P}M) \setminus \{\varnothing\}$ is cancellative if and only if M is trivial.

d) Prove that $(\mathbb{P}S) \setminus \{\varnothing\}$ is a right zero semigroup if and only if S is a right zero semigroup.

✶1.11 Let $X = \{1, \ldots, n\}$ with $n \geqslant 2$. Recalling the cycle notation for permutations from group theory, let $\tau = (1\ 2)$ and $\zeta = (1\ 2\ \ldots\ n-1\ n)$. Note that $\tau, \zeta \in \mathcal{S}_X$; indeed, from elementary group theory, we know

that $\mathcal{S}_X = \langle \tau, \zeta \rangle$. For any $i, j \in X$ with $i \neq j$, let $|i\ j|$ denote the transformation $\varphi_{i,j} \in \mathcal{T}_X$ such that $i\varphi_{i,j} = j\varphi_{i,j} = j$, and $x\varphi_{i,j} = x$ for $x \notin \{i, j\}$.

a) Prove the following four identities when $n \geqslant 3$, and only the last identity for $n \geqslant 2$; note that the elements appearing in first three identities all lie in $\mathcal{T}_X$ only when $n \geqslant 3$:

$$\begin{aligned} (1\ i)|1\ 2|(1\ i) &= |i\ 2| && \text{for } i \geqslant 3; \\ (2\ j)|1\ 2|(2\ j) &= |1\ j| && \text{for } j \geqslant 3; \\ (1\ i)(2\ j)|1\ 2|(2\ j)(1\ i) &= |i\ j| && \text{for } i, j \geqslant 3 \text{ and } i \neq j; \\ (i\ j)|i\ j|(i\ j) &= |j\ i| && \text{for } i, j \geqslant 1 \text{ and } i \neq j. \end{aligned}$$

b) Let $\varphi \in \mathcal{T}_X$. Suppose $|\operatorname{im} \varphi| = r < n$. Let $i, j \in X$ with $i \neq j$ be such that $i\varphi = j\varphi$. Let $k \in X \setminus \operatorname{im} \varphi$. Show that $\varphi = |i\ j|\varphi'$, where $i\varphi' = k$ and $x\varphi' = x\varphi$ for $x \neq i$.

c) Deduce that $\mathcal{T}_X = \langle \tau, \zeta, |1\ 2| \rangle$.

1.12 Let S be a finite monoid. Prove that $x \in S$ is right-invertible if and only if it is left-invertible. [Hint: use the fact that x is periodic.]

∗1.13 Prove than an element of $\mathcal{T}_X$ is

a) left-invertible if and only if it is surjective;

b) right-invertible if and only if it is injective.

1.14 Let $(S, \leqslant)$ be a lattice.

a) Prove that $(x \sqcap y) \sqcup x = x$ and $(x \sqcup y) \sqcap x = x$ for any $x, y \in S$.

b) Deduce that

$$\begin{aligned} &(\forall x, y, z \in S)\big(x \sqcap (y \sqcup z) = (x \sqcap y) \sqcup (x \sqcap z)\big) \\ &\qquad \Leftrightarrow (\forall x, y, z \in S)\big(x \sqcup (y \sqcap z) = (x \sqcup y) \sqcap (x \sqcup z)\big). \end{aligned}$$

[Equivalently: $\sqcap$ distributes over $\sqcup$ if and only if $\sqcup$ distributes over $\sqcap$.]

∗1.15 Give an example of a map φ from a monoid S to a monoid T that is a homomorphism but not a monoid homomorphism.

∗1.16 Let S and T be semigroups and let $\varphi : S \to T$ be a homomorphism. The homomorphism φ is a *categorical monomorphism* if, for any semigroup U and homomorphisms $\psi_1, \psi_2 : U \to S$,

$$\psi_1 \circ \varphi = \psi_2 \circ \varphi \Rightarrow \psi_1 = \psi_2, \tag{1.16}$$

and a *categorical epimorphism* if, for any semigroup U and homomorphisms $\psi_1, \psi_2 : T \to U$,

$$\varphi \circ \psi_1 = \varphi \circ \psi_2 \Rightarrow \psi_1 = \psi_2. \tag{1.17}$$

[These are the definitions of 'monomorphism' and 'epimorphism' used in category theory; the word 'categorical' is simply being used to avoid ambiguity here.]

a) Prove that φ is a monomorphism (as defined on page 19) if and only if it is a categorical monomorphism. [Therefore, for semigroups, monomorphisms and categorical monomorphisms coincide and there is no risk of confusion in using the term 'monomorphism'.]

b) i) Prove that a surjective homomorphism is a categorical epimorphism.

ii) Prove that the inclusion map $\iota : \mathbb{N} \to \mathbb{Z}$ is a categorical epimorphism. [Hint: prove the contrapositive of (1.17) with $\varphi = \iota$.]

[Therefore, for semigroups, categorical epimorphisms are not necessarily surjective. For groups, 'surjective homomorphism' and 'categorical epimorphism' are equivalent. Some authors define 'epimorphism' as 'surjective homomorphism' for semigroups, but this risks confusion.]

1.17 Prove that if we restrict the maps ρ_x in Theorem 1.22 to S (instead of S^1), then the map $x \mapsto \rho_x$ may or may not be injective. [Hint: show that this map is injective if S is a right zero semigroup but not if it is a left zero semigroup.]

1.18 Let Y be a semilattice. Prove that Y is a subdirect product of copies of the two-element semilattice $T = \{e, z\}$, where $e > z$.

1.19 Let I and J be ideals of S such that $I \subseteq J$. Prove that $S/J \simeq (S/I)/(J/I)$.

1.20 Let I and J be ideals of S. Prove that $I \cap J$ and $I \cup J$ are ideals. [Remember to prove that $I \cap J \neq \varnothing$.] Prove that $(I \cup J)/J \simeq I/(I \cap J)$.

1.21 Let S be a semigroup with a zero and let T be a subset of S that contains 0_S and at least one other element. Prove that $T = G \cup \{0_S\}$ for some subgroup G of S if and only if $tT = Tt = T$ for all $t \in T \setminus \{0_S\}$. [This is an analogue of Lemma 1.9 for groups with a zero adjoined.]

NOTES

Most of the definitions and results in this chapter are 'folklore'. ◆ The exposition owes much to the standard accounts in Clifford & Preston, *The Algebraic Theory of Semigroups*, ch. 1 and Howie, *Fundamentals of Semigroup Theory*, ch. 1, which are probably the *ne plus ultra* of how to explain this material, and to a lesser extent Grillet, *Semigroups*, ch. 1 and Higgins, *Techniques of Semigroup Theory*, ch. 1. ◆ The number of non-isomorphic semigroups of order 8 is from Distler, 'Classification and Enumeration of Finite Semigroups', Table A.16. Exercise 1.11 appears as Howie, *Fundamentals of Semigroup Theory*, Exercise 1.6, but contains a minor error in the original. ◆ For an alternative approach to basic semigroup theory, Ljapin, *Semigroups* covers fundamental topics in much greater detail. For an account of structure theory that allows a semigroup to be

empty, see Grillet, *Semigroups*. For further reading on the issues discussed in Exercise 1.16, the standard text on category theory remains Mac Lane, *Categories for the Working Mathematician*. For the situation for groups, see Linderholm, 'A group epimorphism is surjective'.

Free semigroups & presentations

2

> ‘ *how can we think both of presentations as conforming to objects, and objects as conforming to presentations?* is, not the first, but the highest task of transcendental philosophy. ’
>
> — Friedrich Wilhelm Joseph von Schelling, *System of Transcendental Philosophy*, § 3.

Informally, a free semigroup on a set A is the unique biggest, most ‘general’ semigroup generated by [any set in bijection with] A, in the sense that all other semigroups generated by A are homomorphic images (and thus factor semigroups) of the free semigroup on A. This chapter studies some of the interesting properties of free semigroups and then explains their role in semigroup presentations, which can be used to define and manipulate semigroups as factor semigroups of free semigroups.

ALPHABETS AND WORDS

Alphabet, letter, word

An *alphabet* is an abstract set of elements called *letters* or *symbols*. Let A be an alphabet. A *word* over A is a finite sequence $(a_1, a_2, \dots, a_m)$, where each term a_i of the sequence is a letter from A. The *length* of this word is m. There is also a word of length 0, which is the empty sequence $()$. This is called the *empty word*. The set of all words (including the empty word) over A is denoted A^*. The set of all non-empty words (that is, of length 1 or more) over A is denoted A^+.

A^+, A^*

Multiplication of words

Multiplication of words is simply concatenation: that is, for all words $(a_1, a_2, \dots, a_m), (b_1, b_2, \dots, b_n) \in A^*$,

$$(a_1, a_2, \dots, a_m)(b_1, b_2, \dots, b_n) = (a_1, a_2, \dots, a_m, b_1, b_2, \dots, b_n)$$

It is easy to see that this multiplication is associative and so A^* is a semigroup; furthermore, the empty word $()$ is an identity and so A^* is a monoid. Since the product of two words of non-zero length must itself have non-zero length, A^+ is a subsemigroup of A^*; indeed, A^* is [isomorphic to] $(A^+)^1$.

Notation for words

Because of associativity, we simply write $a_1 a_2 \cdots a_n$ for $(a_1, a_2, \dots, a_n)$

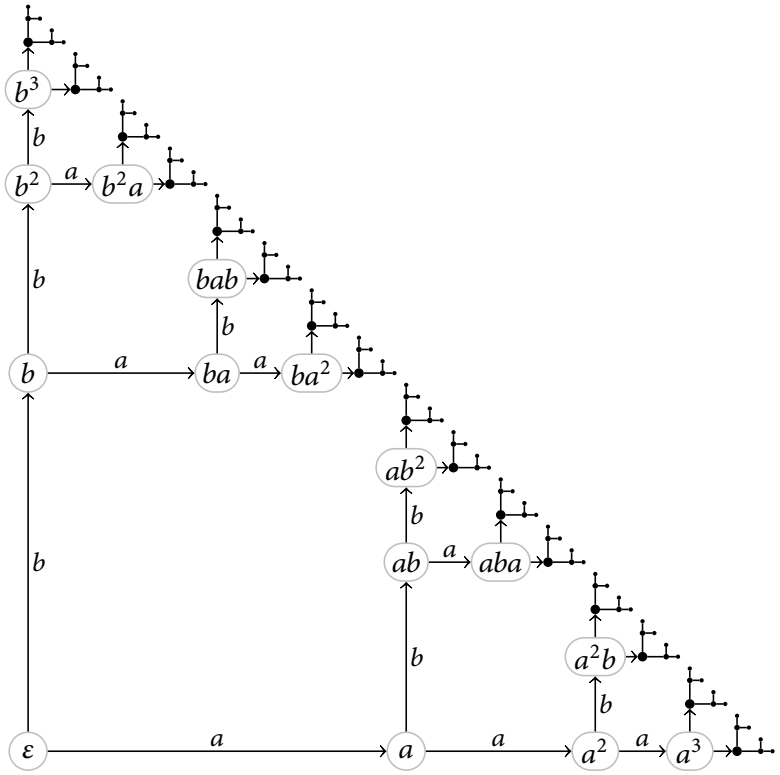


FIGURE 2.1
Part of the Cayley graph $\Gamma(A^*, A)$, where $A = \{a, b\}$.

and write ε for the empty word. For any word $u \in A^*$, denote the length of u by $|u|$, and notice that $|u| = 0$ if and only if $u = \varepsilon$. Note further that $|uv| = |u| + |v|$ for any $u, v \in A^*$.

Subword

A *subword* of a word $a_1 a_2 \cdots a_n$ (where $a_i \in A$) is any word of the form $a_i \cdots a_j$, where $1 \leqslant i \leqslant j \leqslant n$. A *prefix* of $a_1 a_2 \cdots a_n$ is a subword $a_1 \cdots a_i$, where $1 \leqslant i \leqslant n$.

The Cayley graph $\Gamma(A^*, A)$ is an infinite tree; an example for $A = \{a, b\}$ is shown in Figure 2.1. This is obvious, because if we start at ε and follow the path labelled by $u \in A^*$, then we end up at the vertex u. Thus a path uniquely determines a vertex and so the graph must be a tree.

UNIVERSAL PROPERTY

Free semigroup

Let F be a semigroup and let A be an alphabet. Let $\iota : A \hookrightarrow F$ be an embedding of A into F. Then (F, ι) is a *free semigroup on* A if, for any semigroup S and map $\varphi : A \to S$, there is a unique homomorphism $\varphi^+ : F \to S$ that extends φ (that is, with $\iota\varphi^+ = \varphi$). Using

diagrams, this definition says that (F, ι) is a free semigroup on A if

$$\left.\begin{array}{l} \text{for all } \begin{array}{ccc} A & \xhookrightarrow{\iota} & F \\ & \searrow^{\varphi} & \\ & & S \end{array} \text{, there exists a unique} \\ \text{homomorphism } \varphi^+ \text{ such that } \begin{array}{ccc} A & \xhookrightarrow{\iota} & F \\ & \searrow^{\varphi} & \downarrow{\scriptstyle \varphi^+} \\ & & S \end{array}. \end{array}\right\} \quad (2.1)$$

Usually, we just write 'F is a free semigroup on A' instead of the precisely correct '(F, ι) is a free semigroup on A'.

Uniqueness of the free semigroup on A

PROPOSITION 2.1. *Let A be an alphabet and let F be a semigroup. Then F is a free semigroup on A if and only if F is isomorphic to A^+.*

Proof of 2.1. Part 1. Let us first show that A^+ is a free semigroup on A. Let $\iota : A \hookrightarrow A^+$ be the natural embedding map. Let S be a semigroup and $\varphi : A \to S$ be a map. Define $\varphi^+ : A^+ \to S$ by

$$(a_1 a_2 \cdots a_n)\varphi^+ = (a_1\varphi)(a_2\varphi)\cdots(a_n\varphi). \quad (2.2)$$

It is easy to see that φ^+ is a homomorphism and that $\iota\varphi^+ = \varphi$. We now have to prove that φ^+ is unique. So let $\psi : A^+ \to S$ be an arbitrary homomorphism with $\iota\psi = \varphi$. For any $a_1 a_2 \cdots a_n \in A^+$,

$$\begin{aligned} &(a_1 a_2 \cdots a_n)\psi \\ &= (a_1\psi)(a_2\psi)\cdots(a_n\psi) && \text{[since } \psi \text{ is a homomorphism]} \\ &= (a_1\varphi^+)(a_2\varphi^+)\cdots(a_n\varphi^+) && \text{[since } \iota\psi = \varphi = \iota\varphi^+\text{]} \\ &= (a_1 a_2 \cdots a_n)\varphi^+. && \text{[since } \varphi^+ \text{ is a homomorphism]} \end{aligned}$$

and so $\psi = \varphi^+$. Hence φ^+ is the unique homomorphism from A^+ to S with $\iota\varphi^+ = \varphi$, and so A^+ is free on A.

Now suppose that F is isomorphic to A^+ via an isomorphism $\vartheta : A^+ \to F$. The embedding map is $\vartheta\iota : A \hookrightarrow F$. Let $\varphi : A \to S$ be a map. Let $\tau = \vartheta\varphi^+$ (where φ^+ is the homomorphism defined in (2.2)); then $\tau : F \to S$ is a homomorphism extending φ. To see that it is unique, let $\sigma : F \to S$ be an arbitrary homomorphism extending φ. Then $\vartheta^{-1}\sigma : A^+ \to S$ is a homomorphism extending φ. Since A^+ is a free semigroup, $\vartheta^{-1}\sigma = \varphi^+$, and so $\sigma = \mathrm{id}_F\sigma = \vartheta\vartheta^{-1}\sigma = \vartheta\varphi^+ = \tau$. So $\tau : F \to S$ is the unique homomorphism extending φ and so F is a free semigroup on A.

Part 2. Suppose that F is a free semigroup on A; the aim is to show that F is isomorphic to A^+. Let $\iota_1 : A \hookrightarrow A^+$ and $\iota_2 : A \hookrightarrow F$ be the embedding maps. Since A^+ is free on A, we can put ι_1, A^+, ι_2 and F in the places of ι, F, φ and S in (2.1) to see that there is a homomorphism ι_2^+ such that

$$\begin{array}{ccc} A & \xhookrightarrow{\iota_1} & A^+ \\ & \searrow^{\iota_2} & \downarrow{\scriptstyle \iota_2^+} \\ & & F \end{array}. \quad (2.3)$$

Similarly, since F is free on A, we can put ι_2, F, ι_1 and A^+ in the places of ι, F, φ and S in (2.1) to see that there is a homomorphism ι_1^+ such that

$$\begin{array}{ccc} A & \xhookrightarrow{\iota_2} & F \\ & \searrow^{\iota_1} & \downarrow{\scriptstyle \iota_1^+} \\ & & A^+ \end{array}. \tag{2.4}$$

Combining (2.3) and (2.4) in two ways, we get the following diagrams:

$$\begin{array}{ccc} A & \xhookrightarrow{\iota_1} & A^+ \\ & \searrow^{\iota_2} & \downarrow{\scriptstyle \iota_2^+} \\ & & F \\ & \searrow^{\iota_1} & \downarrow{\scriptstyle \iota_1^+} \\ & & A^+ \end{array} \quad \text{and} \quad \begin{array}{ccc} A & \xhookrightarrow{\iota_2} & F \\ & \searrow^{\iota_1} & \downarrow{\scriptstyle \iota_1^+} \\ & & A^+ \\ & \searrow^{\iota_2} & \downarrow{\scriptstyle \iota_2^+} \\ & & F \end{array}. \tag{2.5}$$

Therefore $\iota_1 = \iota_1\iota_2^+\iota_1^+$ and $\iota_2 = \iota_2\iota_1^+\iota_2^+$. In diagrammatic terms, this corresponds to simplifying the diagrams in (2.5) to give

$$\begin{array}{ccc} A & \xhookrightarrow{\iota_1} & A^+ \\ & \searrow^{\iota_1} & \downarrow{\scriptstyle \iota_2^+\iota_1^+} \\ & & A^+ \end{array} \quad \text{and} \quad \begin{array}{ccc} A & \xhookrightarrow{\iota_2} & F \\ & \searrow^{\iota_2} & \downarrow{\scriptstyle \iota_1^+\iota_2^+} \\ & & F \end{array}. \tag{2.6}$$

Clearly the following diagrams commute:

$$\begin{array}{ccc} A & \xhookrightarrow{\iota_1} & A^+ \\ & \searrow^{\iota_1} & \downarrow{\scriptstyle \mathrm{id}_{A^+}} \\ & & A^+ \end{array} \quad \text{and} \quad \begin{array}{ccc} A & \xhookrightarrow{\iota_2} & F \\ & \searrow^{\iota_2} & \downarrow{\scriptstyle \mathrm{id}_F} \\ & & F \end{array}. \tag{2.7}$$

Therefore, by the left-hand diagrams in (2.6) and (2.7), if we put ι_1, A^+, ι_1 and A^+ in place of ι, F, φ, and S in (2.1), then the homomorphisms $\iota_2^+\iota_1^+$ and id_{A^+} are both possibilities for φ^+. But (2.1) requires that there is a unique such homomorphism φ^+, so $\iota_2^+\iota_1^+ = \mathrm{id}_{A^+}$. Similarly, using the right-hand diagrams in (2.6) and (2.7), we obtain $\iota_1^+\iota_2^+ = \mathrm{id}_F$. Hence ι_1^+ and ι_2^+ are mutually inverse isomorphisms, and so F is isomorphic to A^+. 2.1

Free monoids

We could repeat the discussion above, but for monoids instead of semigroups. Let F be a monoid and let A be an alphabet, and let $\iota : A \hookrightarrow F$ be an embedding of A into F. Then (F, ι) is a *free monoid on* A if, for any monoid S and map $\varphi : A \to S$, there is a unique monoid homomorphism $\varphi^* : F \to S$ extending φ; that is, with $\iota\varphi^* = \varphi$. One can prove an analogy of Proposition 2.1 for monoids, showing that a monoid F is a free on A if and only if $F \simeq A^*$. As with free semigroups, we usually write 'F is the free monoid on A' instead of '(F, ι) is the free monoid on A'.

PROPERTIES OF FREE SEMIGROUPS

In preparation for our study of presentations, we begin by examining the structure of free semigroups and monoids.

PROPOSITION 2.2. *Let M be a submonoid of A^*. Let $N = M \setminus \{\varepsilon\}$. Then $N \setminus N^2$ is the unique minimal (monoid) generating set for M.*

Proof of 2.2. Clearly, any generating set for M must contain $N \setminus N^2$. So we must show that $\mathrm{Mon}\langle N \setminus N^2\rangle = M$. Clearly $\mathrm{Mon}\langle N \setminus N^2\rangle \subseteq M$; we have to prove that $M \subseteq \mathrm{Mon}\langle N \setminus N^2\rangle$. We already know that $\varepsilon \in \mathrm{Mon}\langle N \setminus N^2\rangle$, so it remains to show that $N \subseteq \mathrm{Mon}\langle N \setminus N^2\rangle$.

Assume that all words of length less than ℓ in N lie in $\mathrm{Mon}\langle N \setminus N^2\rangle$. Let $u \in N$ with $|u| = \ell$. If $u \in N \setminus N^2$, then $u \in \mathrm{Mon}\langle N \setminus N^2\rangle$. On other other hand, if $u \notin N \setminus N^2$, then $u \in N^2$ and so $u = u'u''$ for $u', u'' \in N$. Hence $|u'| = |u| - |u''|$ and $|u''| = |u| - |u'|$. Since neither u' nor u'' is the empty word, this gives $|u'|, |u''| < |u| = \ell$. So, by assumption, $u', u'' \in \mathrm{Mon}\langle N \setminus N^2\rangle$ and so $u \in \mathrm{Mon}\langle N \setminus N^2\rangle$. Hence, by induction, $N \subseteq \mathrm{Mon}\langle N \setminus N^2\rangle$. 2.2

Base

The *base* of a submonoid or subsemigroup M of A^* is defined to be $N \setminus N^2$, where $N = M \setminus \{\varepsilon\}$. Thus the base is the unique minimal monoid generating set for M if M is a submonoid, and is the unique minimal generating set for M if M is a subsemigroup that is not a submonoid. As an immediate application of Proposition 2.2, we see that A is the base of A^* and A^+.

PROPOSITION 2.3. *A semigroup S is free if and only if every element of S has a unique representative as a product of elements of $S \setminus S^2$.*

Proof of 2.3. Clearly every element of A^+ has a unique representative as a product of elements of $A = A^+ \setminus (A^+)^2$.

So assume that every element of S has a unique representative as a product of elements of $A = S \setminus S^2$. We will show that S satisfies the definition of freedom. Let T be a semigroup and $\varphi : A \to T$ a map. Define a map $\varphi^+ : S \to T$ by letting $s\varphi^+ = (a_1\varphi)(a_2\varphi)\cdots(a_n\varphi)$, where $a_1a_2\cdots a_n$ is the unique representative of s as a product of elements $a_i \in A$. Notice that if $t \in S$ is uniquely represented $b_1\cdots b_m$ where $b_i \in A$, then st has unique representative $a_1\cdots a_nb_1\cdots b_m$. Hence φ^+ is a homomorphism. It is clear that φ^+ is the unique homomorphism extending φ and so S is free on A. 2.3

Free monoid of rank 2 contains a free monoid of countably infinite rank

PROPOSITION 2.4. *Let $A = \{x, y\}$. Let $B = \{b_i : i \in \mathbb{N}\}$. Then A^* contains a submonoid isomorphic to B^*.*

Proof of 2.4. Define a map $\varphi : B \to A^*$ by $b_i\varphi = xy^ix$. Since B^* is free on B, this map φ extends to a unique homomorphism, which we also denote φ, from B^* to A^*.

Suppose, with the aim of obtaining a contradiction, that φ is not injective. Then there exist $u, v \in B^*$ with $u\varphi = v\varphi$.

Suppose u and v begin with the same symbol b; that is, $u = bu'$ and $v = bv'$. Then $(b\varphi)(u'\varphi) = (b\varphi)(v'\varphi)$ and so $u'\varphi = v'\varphi$ by cancellativity in A^*. So we can replace u by u' and v by v' and repeat this process until we have words u and v beginning with different symbols. Therefore assume that u and v begin with symbols b_i and b_j respectively, where $i \neq j$; that is, $u = b_i u'$ and $v = b_j v'$.

Then $xy^i x(u'\varphi) = (b_i\varphi)(u'\varphi) = (b_j\varphi)(v'\varphi) = xy^j x(v'\varphi)$. Assume $i > j$; the other case is similar. By cancellativity in A^*, we have $y^{i-j}x(u'\varphi) = x(v'\varphi)$, which is a contradiction since $i - j > 0$. Therefore φ is injective and so B^* is isomorphic to $\operatorname{im}\varphi$. 2.4

As a consequence of Proposition 2.4, we see that the free monoid on $\{x, y\}$ contains submonoids isomorphic to all free monoids on countable sets. A similar result holds for free semigroups.

Free semigroups can contain non-free subsemigroups

EXAMPLE 2.5. Let $A = \{x\}$ and let $S = \langle x^2, x^3\rangle$. Then $S \setminus S^2 = \{x^2, x^3\}$. But $x^5 \in S$ and $x^5 = x^2x^3 = x^3x^2$, so x^5 has two distinct representatives as a product of elements of $\{x^2, x^3\}$. Hence S is not a free semigroup by Proposition 2.3.

Example 2.5 shows that a free semigroup contains subsemigroups that are not themselves free. In contrast, every subgroup of a free group is itself a free group by the famous Nielsen–Schreier theorem.

SEMIGROUP PRESENTATIONS

Every semigroup is a quotient of a free semigroup

The reason why free semigroups are interesting is that every semigroup is isomorphic to a quotient of a free semigroup. To see this, let $\varphi : A \to S$ be such that $\operatorname{im}\varphi$ generates S. (We could, for instance, choose A to be a set of the same cardinality as S and φ to be a bijection.) Then, φ extends to a homomorphism $\varphi^+ : A^+ \to S$. Since $\operatorname{im}\varphi$ generates S, we have $\operatorname{im}\varphi^+ = S$. By Theorem 1.24, $A^+/\ker\varphi^+ \simeq \operatorname{im}\varphi^+ = S$. That is, S is isomorphic to the quotient $A^+/\ker\varphi^+$.

This is slightly interesting, but its real importance is when we turn it around. Instead of starting with a semigroup and knowing that it is a quotient of a free semigroup, we can specify a free semigroup A^+ and a congruence σ and so *define* the corresponding quotient semigroup A^+/σ.

This is the idea of a semigroup presentation. It allows us to specify and reason about a semigroup as a quotient of a free semigroup: that is, as a quotient A^+/σ for some congruence σ on the free semigroup A^+. By Proposition 2.1, in order to specify the free semigroup, it is sufficient to

specify the alphabet A. In order to specify the congruence σ, it is sufficient to specify some binary relation ρ that generates σ.

Presentations

A *semigroup presentation* is a pair $\mathrm{Sg}\langle A \mid \rho\rangle$, where A is an alphabet and ρ is a binary relation on A^+. The elements of A are called *generating symbols*, and the elements of ρ (which are pairs of words in A^+) are called *defining relations*. The presentation $\mathrm{Sg}\langle A \mid \rho\rangle$ *defines*, or *presents*, any semigroup isomorphic to $A^+/\rho^\#$.

Let S be a semigroup presented by $\mathrm{Sg}\langle A \mid \rho\rangle$. Then S is isomorphic to $A^+/\rho^\#$ and so there is a one-to-one correspondence between elements of S and $\rho^\#$-classes. Thus we can think of a word $w \in A^+$ as *representing* the element of S corresponding to $[w]_{\rho^\#}$. If $u, v \in A^+$ represent the same element of S (that is, if $(u, v) \in \rho^\#$, or, equivalently, if $[u]_{\rho^\#} = [v]_{\rho^\#}$), we say that u and v are *equal in* S and write $u =_S v$.

Assignment of generators

Let T be a semigroup. Let $\varphi : A \to T$ be a map such that $A\varphi$ generates T; such a map is called an *assignment of generators*. In this case, the unique homomorphism $\varphi^+ : A^+ \to T$ extending φ is surjective.

Satisfying a defining relation

The semigroup T *satisfies* a defining relation $(u, v) \in \rho$ with respect to an assignment of generators $\varphi : A \to T$ if $u\varphi^+ = v\varphi^+$. Notice that T satisfies all defining relations in ρ with respect to $\varphi : A \to T$ if and only if $\rho \subseteq \ker \varphi^+$. By definition, any semigroup defined by the presentation $\mathrm{Sg}\langle A \mid \rho\rangle$ satisfies the defining relations ρ with respect to the assignment of generators $(\rho^\#)^\natural|_A : A \to A^+/\rho^\#$.

PROPOSITION 2.6. *Let T be a semigroup, and suppose T satisfies the defining relations in ρ with respect to an assignment of generators $\varphi : A \to T$. Then T is a homomorphic image of the semigroup presented by* $\mathrm{Sg}\langle A \mid \rho\rangle$.

Proof of 2.6. Since T satisfies the defining relations ρ with respect to φ, we have $\rho \subseteq \ker \varphi^+$. Since $\ker \varphi^+$ is a congruence by Theorem 1.24, and $\rho^\#$ is the smallest congruence containing ρ, it follows that $\rho^\# \subseteq \ker \varphi^+$. So the map $\psi : A^+/\rho^\# \to T$ defined by $[u]_{\rho^\#}\psi = u\varphi^+$ is a well-defined homomorphism, and is clearly surjective since φ^+ is surjective. 2.6

By Proposition 2.6, we can think of semigroup presented by $\mathrm{Sg}\langle A \mid \rho\rangle$ as the largest semigroup generated by A and satisfying the defining relations in ρ.

Elementary transition

An *elementary ρ-transition* is a pair $(w, w') \in (\rho^{\mathsf{C}})^{\mathsf{S}} = \rho^{\mathsf{C}} \cup (\rho^{\mathsf{C}})^{-1}$, which we denote $w \leftrightarrow_\rho w'$. By Proposition 1.27, $w \leftrightarrow_\rho w'$ if and only if w' can be obtained from w by substituting a subword y for a subword x of w, where $(x, y) \in \rho$ or $(y, x) \in \rho$. In this situation, we say that we *apply* the defining relation (x, y) or (y, x) to the word w and obtain w'.

Let $u, v \in A^+$. If there is a sequence of elementary ρ-transitions $u = w_0 \leftrightarrow_\rho \dots \leftrightarrow_\rho w_n = v$, then we say (u, v) is a *consequence* of ρ, or (u, v) can be *deduced* from ρ. The following result shows the connection between this notion and presentations:

PROPOSITION 2.7. *Let S be presented by $\mathrm{Sg}\langle A \mid \rho\rangle$ and let $u, v \in A^+$. Then $u =_S v$ if and only if (u, v) is a consequence of ρ; that is, if and only if there is a sequence of elementary ρ-transitions*

$$u = w_0 \leftrightarrow_\rho w_1 \leftrightarrow_\rho \ldots \leftrightarrow_\rho w_n = v.$$

Proof of 2.7. First of all, note that

$$\begin{aligned}
& u =_S v \\
\Leftrightarrow\ & (u, v) \in \rho^\# \\
\Leftrightarrow\ & (u, v) \in (\rho^{\mathsf{C}})^{\mathsf{E}} && \text{[by Proposition 1.29]} \\
\Leftrightarrow\ & (u, v) \in \mathrm{id}_{A^+} \cup \textstyle\bigcup_{n=1}^{\infty} (\rho^{\mathsf{C}} \cup (\rho^{\mathsf{C}})^{-1})^n && \text{[by Proposition 1.26(f)]} \\
\Leftrightarrow\ & (u = v) \vee (\exists n \in \mathbb{N})\big((u, v) \in (\rho^{\mathsf{C}} \cup (\rho^{\mathsf{C}})^{-1})^n\big) \\
\Leftrightarrow\ & (\exists n \in \mathbb{N} \cup \{0\})(\exists w_0, \ldots, w_n \in A^+) \\
& \qquad \big[(u = w_0) \wedge (w_n = v) \\
& \qquad\qquad \wedge (\forall i)\big((w_i, w_{i+1}) \in \rho^{\mathsf{C}} \cup (\rho^{\mathsf{C}})^{-1}\big)\big] \\
\Leftrightarrow\ & (\exists n \in \mathbb{N} \cup \{0\})(\exists w_0, \ldots, w_n \in A^+) \\
& \qquad \big[(u = w_0) \wedge (w_n = v) \wedge (\forall i)\big(w_i \leftrightarrow_\rho w_{i+1}\big)\big)\big].
\end{aligned}$$

Hence $u =_S v$ if and only if there is there is a sequence of elementary ρ-transitions from u to v. 2.7

The next result gives a usable condition for when a given presentation defines a particular semigroup. Afterwards, we will see how this result yields a practical proof method.

Condition for $\mathrm{Sg}\langle A \mid \rho\rangle$ to define S

PROPOSITION 2.8. *Let S be a semigroup. Then $\mathrm{Sg}\langle A \mid \rho\rangle$ presents S if and only if there is an assignment of generators $\varphi : A \to S$ such that*

a) *S satisfies the defining relations in ρ with respect to φ, and*

b) *if $u, v \in A^+$ are such that $u\varphi^+ = v\varphi^+$, then (u, v) is a consequence of ρ.*

Proof of 2.8. Suppose first that $\mathrm{Sg}\langle A \mid \rho\rangle$ presents S. Then S is isomorphic to $A^+/\rho^\#$, so we can let φ be the restriction of natural homomorphism $(\rho^\#)^\natural|_A : A^+ \to A^+/\rho^\#$. Then condition a) holds from the definition and condition b) holds from Proposition 2.7.

Now suppose that conditions a) and b) hold. Since S satisfies the defining relations in ρ, we have $\rho \subseteq \ker \varphi^+$ and so $\rho^\# \subseteq \ker \varphi^+$. If $(u, v) \in \ker \varphi^+$, then $u\varphi^+ = v\varphi^+$ and so (u, v) is a consequence of ρ and hence $(u, v) \in \rho^\#$ by Proposition 2.7. Hence $\rho^\# = \ker \varphi^+$ and therefore $S \simeq A^+/\ker \varphi^+ \simeq A^+/\rho^\#$; thus $\mathrm{Sg}\langle A \mid \rho\rangle$ presents S. 2.8

There is a standard three-step method for directly proving that a presentation defines a particular semigroup:

Method for proving $\mathrm{Sg}\langle A \mid \rho\rangle$ defines S

METHOD 2.9. To prove that a presentation $\mathrm{Sg}\langle A \mid \rho\rangle$ defines a particular semigroup S:

1) Define an assignment of generators $\varphi : A \to S$, and prove that S satisfies the defining relations in ρ with respect to φ.
2) Find a set of words $N \subseteq A^+$ such that for every word $w \in A^+$ there is a word $\widehat{w} \in N$ such that $(w, \widehat{w})$ is a consequence of ρ.
3) Prove that $\varphi^+|_N$ is injective.

In Method 2.9, step 1 establishes that condition a) of Proposition 2.8 holds. Now let $u, v \in A^+$ be such that $u\varphi^+ = v\varphi^+$. Step 2 shows that $(u, \widehat{u})$ and $(v, \widehat{v})$ are consequences of ρ; that is, there are sequences of elementary ρ-transitions $u \leftrightarrow_\rho \ldots \leftrightarrow_\rho \widehat{u}$ and $v \leftrightarrow_\rho \ldots \leftrightarrow_\rho \widehat{v}$. Since S satisfies the relations in ρ, this implies that $\widehat{u}\varphi^+ = \widehat{v}\varphi^+$, so step 3 shows that $\widehat{u} = \widehat{v}$, and thus there is a sequence of elementary ρ-transitions $u \leftrightarrow_\rho \ldots \leftrightarrow_\rho \widehat{u} = \widehat{v} \leftrightarrow_\rho \ldots \leftrightarrow_\rho v$; that is, (u, v) is a consequence of ρ. This establishes condition b) of Proposition 2.8 and so proves that $\mathrm{Sg}\langle A \mid \rho\rangle$ presents S.

Before giving some examples to illustrate the theory described above, we introduce a convention to simplify notation. When we explicitly list generating symbols and defining relations in a presentation, we do not write the braces { } enclosing the list of elements in the two sets. So instead of $\mathrm{Sg}\langle\{a_1, a_2, \ldots\} \mid \{(u_1, v_1), (u_2, v_2), \ldots\}\rangle$, we write $\mathrm{Sg}\langle a_1, a_2, \ldots \mid (u_1, v_1), (u_2, v_2), \ldots\rangle$.

EXAMPLE 2.10. a) Let us prove that the presentation $\mathrm{Sg}\langle A \mid \rangle$ (with no defining relations) defines the free semigroup A^+. To see this, it suffices to notice that $\varnothing^\# = \mathrm{id}_{A^+}$, and $A^+/\mathrm{id}_{A^+} \simeq A^+$.

[Following Method 2.9 for the sake of illustration, let φ be the embedding map $\iota : A \hookrightarrow A^+$. Clearly A^+ trivially satisfies all defining relations with respect to φ; this is step 1. Let $N = A^+$; then every word in A^+ itself in N and so step 2 is immediately proved. Finally, step 3 is trivial since φ^+ is the identity map and so injective.]

b) Now we prove that the presentation $\mathrm{Sg}\langle a \mid (a^2, a)\rangle$ defines the trivial semigroup $\{e\}$. To see this, it suffices to notice that $\{(a^2, a)\}^\# = \{a\}^+ \times \{a\}^+$.

[Following Method 2.9 for the sake of illustration, let $\varphi : \{a\} \to \{e\}$ be given by $a\varphi = e$. Then $a^2\varphi^+ = (a\varphi^+)^2 = e^2 = e = a\varphi^+$ and so the semigroup $\{e\}$ satisfies the defining relation (step 1). Let $N = \{a\}$. Then any word in $\{a\}^+$ can be transformed to the unique word $a \in N$ by repeatedly applying the defining relation (step 2). Finally, N contains only a single element and hence $\varphi^+|_N$ is trivially injective (step 3).]

c) Less trivially, we now prove that the presentation $\mathrm{Sg}\langle A \mid (ab, a) : a, b \in A\rangle$ defines a left zero semigroup on a set of size $|A|$. Following Method 2.9, let S be the left zero semigroup with $|A|$ elements and let $\varphi : A \to S$ be a bijection. Then $(ab)\varphi^+ = (a\varphi^+)(b\varphi^+) = a\varphi^+$ since S is a left zero semigroup, and so S satisfies the defining relations (step 1). Let $N = A$; then any word in A^+ can be transformed to one in N by

applying defining relations to replace a subword ab by a; this yields a shorter word and so ends with a word in A (step 2). Finally, if $a, b \in N$ are such that $a\varphi^+ = b\varphi^+$, then $a\varphi = a\varphi^+ = b\varphi^+ = b\varphi$, which implies $a = b$ since φ is a bijection; thus $\varphi^+|_N$ is injective (step 3).

d) Consider the set $M_3(\mathbb{Z})$ of all 3×3 integer matrices. Let

$$P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}, \quad Q = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad R = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

and let S be the subsemigroup of $M_3(\mathbb{Z})$ generated by $\{P, Q, R\}$. Let us prove that S is presented by

$$\mathrm{Sg}\langle a, b, c \mid (ba, abc), (ca, ac), (cb, bc)\rangle.$$

First, let $\varphi : \{a, b, c\} \to S$ be given by $a\varphi = P$, $b\varphi = Q$, and $c\varphi = R$. Straightforward calculations show that S satisfies the defining relations with respect to φ^+ (step 1). Let

$$N = \{\, a^i b^j c^k : i, j, k \in \mathbb{N} \cup \{0\} \wedge i, j, k \text{ not all } 0 \,\}.$$

Every word in $\{a, b, c\}^+$ can be transformed to one in N as follows: First, by applying the second and third defining relations from left to right, we move all symbols c to the right of the word. Then, if there is some symbol b to the left of a symbol a, we apply the first defining relation, and move the 'new' symbol c to the right of the word. We repeat this step until there is no symbol b to the left of a symbol a. This process must terminate because no application of a relation changes the number of symbols a or b in the word. At the end of the process, we are left with a word in N (step 2). Finally, a simple calculation shows that

$$(a^i b^j c^k)\varphi^+ = \begin{bmatrix} 1 & j & k \\ 0 & 1 & i \\ 0 & 0 & 1 \end{bmatrix},$$

and so if $a^i b^j c^k =_S a^{i'} b^{j'} c^{k'}$, then $i = i'$, $j = j'$, and $k = k'$; hence $\varphi^+|_N$ is injective (step 3). [Note that the *subgroup* of $M_3(\mathbb{Z})$ generated by $\{P, Q, R\}$ is the famous *discrete Heisenberg group* $H_3(\mathbb{Z})$.]

Monoid presentations

We could repeat the discussion of presentations above, but reasoning about monoids instead of semigroups. Every monoid is a quotient of a free monoid. In a monoid presentation $\mathrm{Mon}\langle A \mid \rho\rangle$ the defining relations in ρ are of the form (u, v) for $u, v \in A^*$. In particular, they can be of the form (u, ε) or (ε, u) or $(\varepsilon, \varepsilon)$. The presentation $\mathrm{Mon}\langle A \mid \rho\rangle$ presents the monoid $A^*/\rho^\#$. The notion of an assignment of generators carries over to monoids. The analogies of Propositions 2.6, 2.7, and 2.8 all hold for

monoids, using A^* instead of A^+ and φ^* instead of φ^+ as appropriate. Thus the monoid presented by $\mathrm{Mon}\langle A \mid \rho\rangle$ is the largest monoid generated by A and satisfying the defining relations in ρ. If M is presented by $\mathrm{Mon}\langle A \mid \rho\rangle$, then for $u, v \in A^*$, we have $u =_M v$ if and only if there is a sequence of elementary ρ-transitions from u to v. Finally, Method 2.9 works for monoids, again with φ^* instead of φ^+.

EXAMPLE 2.11. a) Let us prove that the monoid $(\mathbb{N} \cup \{0\}) \times (\mathbb{N} \cup \{0\})$ is presented by $\mathrm{Mon}\langle a, b \mid (ab, ba)\rangle$. Following the monoid version of Method 2.9, let $\varphi : \{a, b\} \to (\mathbb{N} \cup \{0\}) \times (\mathbb{N} \cup \{0\})$ be defined by $a\varphi = (1, 0)$ and $b\varphi = (0, 1)$. Then $(ab)\varphi^* = (1, 0)(0, 1) = (1, 1) = (0, 1)(1, 0) = (ba)\varphi^*$, so $(\mathbb{N} \cup \{0\}) \times (\mathbb{N} \cup \{0\})$ satisfies the defining relations with respect to φ (step 1). Let $N = \{ a^i b^j : i, j \in \mathbb{N} \cup \{0\} \}$ Then every word in $\{a, b\}^*$ can be transformed to one in N by applying the defining relation to move symbols a to the left of symbols b (step 2). Finally, note that if $a^i b^j \varphi^* = a^{i'} b^{j'} \varphi^*$, then $(i, j) = (i', j')$ and so $i = i'$ and $j = j'$; thus $\varphi^*|_N$ is injective (step 3).

b) Let $\tau, \sigma \in \mathcal{T}_{\mathbb{N}}$ be given by

$$\tau = \begin{pmatrix} 1 & 2 & 3 & 4 & \dots \\ 2 & 3 & 4 & 5 & \dots \end{pmatrix}, \qquad \sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & \dots \\ 1 & 1 & 2 & 3 & \dots \end{pmatrix},$$

and let B be the submonoid generated by $\{\tau, \sigma\}$. Let us prove that B is presented by $\mathrm{Mon}\langle b, c \mid (bc, \varepsilon)\rangle$. Define $\varphi : \{b, c\} \to B$ by $b\varphi = \tau$ and $c\varphi = \sigma$. Then $(bc)\varphi^* = \tau\sigma = \mathrm{id}_{\mathbb{N}} = \varepsilon\varphi^*$, so B satisfies the defining relation with respect to φ (step 1). Let $N = \{ c^i b^j : i, j \in \mathbb{N} \cup \{0\} \}$; then every word in $\{b, c\}^*$ can be transformed to one in N by using the defining relations to replace subwords bc by ε (effectively 'deleting' the subword bc) and ultimately yielding one in N (step 2). Finally,

$$\begin{aligned} &(c^i b^j)\varphi^* \\ &= \begin{pmatrix} 1 & 2 & \dots & i+1 & i+2 & \dots \\ 1 & 1 & \dots & 1 & 2 & \dots \end{pmatrix} \\ &\qquad \begin{pmatrix} 1 & 2 & 3 & 4 & \dots \\ j+1 & j+2 & j+3 & j+4 & \dots \end{pmatrix} \\ &= \begin{pmatrix} 1 & 2 & \dots & i+1 & i+2 & \dots \\ j+1 & j+1 & \dots & j+1 & j+2 & \dots \end{pmatrix}, \end{aligned}$$

and so in the image of $c^i b^j$, the image of 1 is $j + 1$ and the maximum element of the domain with image $j + 1$ is $i + 1$. That is, the image determines i and j. Hence $\varphi^*|_N$ is injective (step 3).

This monoid B defined by $\mathrm{Mon}\langle b, c \mid (bc, \varepsilon)\rangle$ is the *bicyclic monoid*. Every element of the bicyclic monoid is represented by a unique word of the form $c^i b^j$, where $i, j \in \mathbb{N} \cup \{0\}$, and we normally work with these representatives of elements of B. Multiplying using these

Bicyclic monoid

representatives is concatenation followed by deletion of subwords bc. That is,

$$c^i b^j \; c^k b^\ell =_B \begin{cases} c^{i+k-j} b^\ell & \text{if } k \geqslant j, \\ c^i b^{j-k+\ell} & \text{if } k \leqslant j. \end{cases}$$

Finite presentation

A presentation is *finite* if both A and ρ are finite. The semigroup is *finitely presented* if is defined by some finite presentation.

Finite presentability is independent of the generating set

PROPOSITION 2.12. *Suppose S is finitely presented and let $\varphi : A \to S$ be an assignment of generators (with A possibly being infinite). Then there exists a finite subset B of A and $\rho \subseteq B^+ \times B^+$ such that $\mathrm{Sg}\langle B \mid \rho\rangle$ is a finite presentation defining S.*

Proof of 2.12. Since S is finitely presented, it is defined by a finite presentation $\mathrm{Sg}\langle C \mid \tau\rangle$. For brevity, let $\psi : C \to S$ be the natural assignment of generators $(\tau^\#)^\natural|_C$, so that $c\psi = [c]_{\tau^\#}$.

For each $c \in C$, there exists a word $c\zeta \in A^+$ such that c and $c\zeta$ represent the same element of S. (We can choose $c\zeta$ to be any word in $(c\psi)(\varphi^+)^{-1}$.) Thus we have a map $\zeta : C \to A^+$ such that $\zeta^+\varphi^+ = \psi^+$. Let

$$B = \{\, b \in A : (\exists c \in C)(b \text{ appears in } c\zeta) \,\};$$

thus $C\zeta \subseteq B^+$ and so $\zeta^+\varphi|_B^+ = \psi^+$. Notice that B is finite since C is finite. Notice that $\langle B\varphi|_B\rangle \supseteq \langle C\psi\rangle = S$, so $\varphi|_B : B \to S$ is an assignment of generators.

Similarly, for every $b \in B$, there exists a word $b\eta \in C^+$ such that b and $b\eta$ represent the same element of S. (We can choose $b\eta$ to be any word in $(b\varphi|_B)(\psi^+)^{-1}$.) Thus we have a map $\eta : B \to C^+$ such that $\eta^+\psi^+ = \varphi|_B^+$. (Figure 2.2 shows the relationship between $\varphi|_B^+$, ψ^+, ζ^+, and η^+.)

Let

$$\rho = \{\, (p\zeta^+, q\zeta^+) : (p,q) \in \tau \,\} \cup \{\, (b, b\eta^+\zeta^+) : b \in B \,\}.$$

Note first that $\rho \subseteq B^+ \times B^+$. Now, if $(p,q) \in \tau$, then S satisfies this defining relation with respect to ψ, so $p\psi^+ = q\psi^+$, and hence $p\zeta^+\varphi|_B^+ = q\zeta^+\varphi|_B^+$. Furthermore, if $b \in B$, then $b\varphi|_B^+ = b\eta^+\psi^+ = b\eta^+\zeta^+\varphi|_B^+$. So S satisfies every defining relation in ρ with respect to $\varphi|_B$.

It remains to prove that if $u, v \in B^+$ are such that $u\varphi|_B^+ = v\varphi|_B^+$, then (u,v) is a consequence of ρ. So suppose $u\varphi|_B^+ = v\varphi|_B^+$. Then $u\eta^+\psi^+ = v\eta^+\psi^+$. So $(u\eta^+, v\eta^+)$ is a consequence of τ. That is, there is a sequence of elementary τ-transitions

$$u\eta^+ = w_0 \leftrightarrow_\tau w_1 \leftrightarrow_\tau \ldots \leftrightarrow_\tau w_n = v\eta^+.$$

So, applying ζ^+ to this sequence, we see that by the definition of ρ, there is a sequence of elementary ρ-transitions

$$u\eta^+\zeta^+ = w_0\zeta^+ \leftrightarrow_\rho w_1\zeta^+ \leftrightarrow_\rho \ldots \leftrightarrow_\rho w_n\zeta^+ = v\eta^+\zeta^+. \tag{2.8}$$

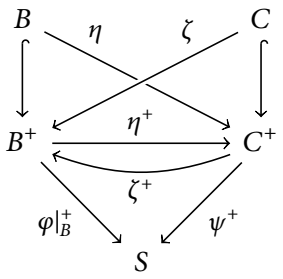


FIGURE 2.2
Maps used in the proof of Proposition 2.12

Suppose $u = u_1u_2 \cdots u_k$ and $v = v_1v_2 \cdots v_\ell$, where $u_i, v_i \in B$. By the definition of ρ, there are also sequences of elementary ρ-transitions

$$\left.\begin{aligned} u = u_1u_2 \cdots u_k &\leftrightarrow_\rho (u_1\eta^+\psi^+)u_2 \cdots u_k \leftrightarrow_\rho \cdots \\ &\leftrightarrow_\rho (u_1\eta^+\psi^+)(u_2\eta^+\psi^+) \cdots (u_k\eta^+\psi^+) = u\eta^+\zeta^+ \end{aligned}\right\} \quad (2.9)$$

and

$$\left.\begin{aligned} v\eta^+\zeta^+ = (v_1\eta^+\psi^+)(v_2\eta^+\psi^+) \cdots (v_\ell\eta^+\psi^+) &\leftrightarrow_\rho \cdots \\ &\leftrightarrow_\rho (v_1\eta^+\psi^+)v_2 \cdots v_\ell \leftrightarrow_\rho v_1v_2 \cdots v_\ell = v. \end{aligned}\right\} \quad (2.10)$$

Concatenating the sequences (2.8), (2.9), and (2.10) shows that (u, v) is a consequence of ρ and so completes the proof. 2.12

We now give two more important examples. Example 2.13 shows that a semigroup can be finitely generated but not finitely presented. Example 2.14 then shows that cancellativity is not a sufficient condition for group-embeddability.

Finitely generated but not finitely presented

EXAMPLE 2.13. Let $X = \{xyz, yz, yt, xy, zy, zyt\} \subseteq \{x, y, z, t\}^+$. Let S be the subsemigroup of $\{x, y, z, t\}^+$ generated by X.

Suppose, with the aim of obtaining a contradiction, that S is finitely presented. Let $A = \{a, b, c, d, e, f\}$ and let $\varphi : A \to S$ be given by

$$\begin{aligned} a\varphi &= xyz, & b\varphi &= yz, & c\varphi &= yt, \\ d\varphi &= xy, & e\varphi &= zy, & f\varphi &= zyt. \end{aligned}$$

Clearly S is presented by $\mathrm{Sg}\langle A \mid \ker\varphi^+\rangle$, since $A^+/\ker\varphi^+ \simeq S$ by Theorem 1.24. Thus, by Proposition 2.12, S is defined by a finite presentation $\mathrm{Sg}\langle A \mid \sigma\rangle$. Assume without loss of generality that σ contains no defining relations of the form (u, u).

Let α be greater than the maximum length of a side of a defining relation in σ. Now,

$$(ab^\alpha c)\varphi^+ = x(yz)^{\alpha+1}yt = (de^\alpha f)\varphi^+.$$

That is, $ab^\alpha c =_S de^\alpha f$. By Proposition 2.7, there is a sequence of elementary σ-transitions

$$ab^\alpha c \leftrightarrow_\sigma \cdots \leftrightarrow_\sigma de^\alpha f. \quad (2.11)$$

For any $\beta \in \mathbb{N} \cup \{0\}$, the word ab^β is the unique word over A representing $(ab^\beta)\varphi = x(yz)^{\beta+1}$, the word $b^\beta c$ is the unique word over A representing $(b^\beta c)\varphi = (yz)^\beta yt$, and for $\beta \neq 0$ the word b^β is the unique word over A representing $(b^\beta)\varphi = (yz)^\beta$. Hence σ cannot contain any defining relation of the form (ab^β, u) or $(b^\beta c, v)$ or (b^β, w). Thus in the sequence of elementary σ-transitions (2.11), the first step must involve applying a defining relation of which one side is $ab^\alpha c$. This contradicts the fact that α is greater than the maximum length of a side of a defining relation in σ. Therefore S is not finitely presented.

Example 2.5 showed that it is possible for a free semigroup to contain subsemigroups that are not themselves free. By showing that a free semigroup can contain finitely generated subsemigroups that are not even finitely presented, Example 2.13 provides an even stronger contrast to the Nielsen–Schreier theorem.

Cancellative but not group-embeddable

EXAMPLE 2.14. Let S be the semigroup presented by $\mathrm{Sg}\langle A \mid \rho\rangle$, where $A = \{a, b, c, d, e, f, g, h\}$ and let $\rho = \{(ae, bf), (cf, de), (dg, ch)\}$. We will prove that S is cancellative but not group-embeddable.

Proving that S is cancellative involves many cases, so we prove the left-cancellativity condition for the generator represented by c; the other cases are similar. Suppose that $cu =_S cv$; we aim to prove that $u =_S v$. Then there is a sequence of elementary ρ-transitions

$$cu = w_0 \leftrightarrow_\rho \ldots \leftrightarrow_\rho w_n = cv. \tag{2.12}$$

Without loss of generality, assume that n is minimal among all such sequences. Suppose, with the aim of obtaining a contradiction, that at some step in this sequence the initial symbol c is altered. This must involve applying one of the defining relations (cf, de) or (dg, ch). Assume the former; the latter case is similar. Thus (2.12) is of the form

$$cu = w_0 \leftrightarrow_\rho \ldots \leftrightarrow_\rho cfw' \leftrightarrow_\rho dew' \leftrightarrow_\rho \ldots \leftrightarrow_\rho w_n = cv.$$

Now, no defining relation has one side starting with a symbol e, so the symbol e must remain in the terms of the sequence until the defining relation (cf, de) is applied again to alter the initial symbol d. (We know that this relation must be applied because the sequence of elementary ρ-transitions ends with $w_n = cv$.) Thus (2.12) is of the form

$$\begin{aligned} cu = w_0 \leftrightarrow_\rho \ldots \leftrightarrow_\rho cfw' \leftrightarrow_\rho dew' \leftrightarrow_\rho \ldots \\ \leftrightarrow_\rho dew'' \leftrightarrow_\rho cfw'' \leftrightarrow_\rho \ldots \leftrightarrow_\rho w_n = cv. \end{aligned}$$

Since the distinguished symbol e is present throughout the subsequence $dew' \leftrightarrow_\rho \ldots \leftrightarrow_\rho dew''$, so does the symbol d. Because the symbols de are not involved in any of the intermediate steps, there is no need to include the two elementary ρ-transitions $cfw' \leftrightarrow_\rho dew'$ and $dew'' \leftrightarrow_\rho cfw''$. That is, we can remove the elementary ρ-transitions $cfw' \leftrightarrow_\rho dew'$ and $dew'' \leftrightarrow_\rho cfw''$ and replace the prefixes de by cf in the subsequence $dew' \leftrightarrow_\rho \ldots \leftrightarrow_\rho dew''$ and obtain a strictly shorter sequence of elementary ρ-transitions from cu to cv. This contradicts the minimality of n. Therefore the initial symbol c is never altered. Thus we can delete the initial symbol c from each step in (2.12) to obtain a sequence of elementary ρ-transitions from u to v. Hence $u =_S v$.

This argument proves that the left-cancellativity condition holds for the generator c. Reasoning similarly for the symbols in $A \setminus \{c\}$ shows

that S is left-cancellative; symmetrical arguments show that S is right cancellative. Thus S is cancellative.

Suppose S is group-embeddable. Then there is a monomorphism $\varphi : S \to G$, where G is a group. Then

$$\begin{aligned}
(ag)\varphi &= (a\varphi)(g\varphi) \\
&= (a\varphi)(e\varphi)(e\varphi)^{-1}(g\varphi) \\
&= (b\varphi)(f\varphi)(e\varphi)^{-1}(g\varphi) && [\text{since } ae =_S bf] \\
&= (b\varphi)(c\varphi)^{-1}(c\varphi)(f\varphi)(e\varphi)^{-1}(g\varphi) \\
&= (b\varphi)(c\varphi)^{-1}(d\varphi)(e\varphi)(e\varphi)(g\varphi) && [\text{since } cf =_S de] \\
&= (b\varphi)(c\varphi)^{-1}(d\varphi)(g\varphi) \\
&= (b\varphi)(c\varphi)^{-1}(c\varphi)(h\varphi) && [\text{since } dg =_S ch] \\
&= (b\varphi)(h\varphi) \\
&= (bh)\varphi.
\end{aligned}$$

But $ag \neq_S bh$, since there is no sequence of elementary ρ-transitions from ag to bh because ag does not contain a subword that forms one side of a defining relation in ρ. This contradicts φ being a monomorphism and so S is not group-embeddable.



Several of the syntactic arguments used in this chapter and in the exercises could be simplified by using the tools of string-rewriting. However, presenting the necessary theory is beyond the scope of this course.

EXERCISES

[See pages 209–215 for the solutions.]

✶2.1 A semigroup S is *equidivisible* if for all $x, y, z, t \in S$, the following holds: Equidivisibility

$$\begin{aligned}
xy = zt \Rightarrow (\exists p \in S)(x = zp \wedge t = py) \\
\vee (\exists q \in S)(z = xq \wedge y = qt).
\end{aligned}$$

a) Prove that groups are equidivisible.

b) Prove that free monoids are equidivisible.

✶2.2 Let $u, v \in A^+$. Prove that

$$uv = vu \Leftrightarrow (\exists w \in A^+)(\exists i, j \in \mathbb{N})(u = w^i \wedge v = w^j).$$

[Hint: to prove the left-hand side implies the right-hand side, use induction on $|uv|$.]

2.3 Let $u, v, w \in A^+$ be such that $uv = vw$.

a) Using induction on $|v|$, prove that there exist $s, t \in A^*$ and $k \in \mathbb{N} \cup \{0\}$ such that $u = st$, $v = (st)^k s$, and $w = ts$.

b) Prove part a) in a different way by letting k be maximal (possibly $k = 0$) such that $v = u^k s$ for some $s \in A^*$.

2.4 Let $u, v \in A^+$. Show that the subsemigroup $\langle u, v \rangle$ is free if and only if $uv \neq vu$.

2.5 Let S be a semigroup and let X be a generating set for S, with $|X| \geqslant 2$. Suppose that for all $x_i, y_i \in X$ and $n \in \mathbb{N}$, we have $x_1 \cdots x_n = y_1 \cdots y_n \Rightarrow (\forall i \in \{1, \ldots, n\})(x_i = y_i)$. Prove that S is free with basis X.

✶2.6 Let $n \in \mathbb{N}$. Let $X = \{x_1, x_2, \ldots, x_n\}$. Let M be the set $\mathbb{P}X$ under the operation of union; then M is a monoid with identity $\varnothing$. The aim of this exercise is to use Method 2.9 to prove that M is defined by $\mathrm{Mon}\langle A \mid \rho \rangle$, where $A = \{a_1, \ldots, a_n\}$ and $\rho = \{ (a_i^2, a_i), (a_i a_j, a_j a_i) : i, j \in \{1, \ldots, n\} \}$.

a) Do step 1 of Method 2.9: define an assignment of generators $\varphi : A \to M$ and show that M satisfies the defining relations in ρ with respect to φ. [Hint: the monoid M is generated by elements $\{x_1\}, \{x_2\}, \ldots, \{x_n\}$.]

b) Do step 2 of Method 2.9: let $N = \{ a_1^{e_1} a_2^{e_2} \cdots a_n^{e_n} : e_i \leqslant 1 \}$ and prove that for every $w \in A^*$ there is a word $\widehat{w} \in N$ such that $(w, \widehat{w})$ is a consequence of ρ.

c) Do step 3 of Method 2.9: prove that $\varphi^*|_N$ is injective.

✶2.7 Prove that $\mathrm{Mon}\langle a, b \mid (aba, \varepsilon) \rangle$ defines $(\mathbb{Z}, +)$.

2.8 Let M be defined by $\mathrm{Mon}\langle A \mid \rho \rangle$, where $A = \{a, b, c\}$ and $\rho = \{(abc, \varepsilon)\}$. Let $N = A^* \setminus A^* abc A^*$, so that N consists of all words over A that do not contain a subword abc. Prove that every element of M has a unique representative in N, and that this representative can be obtained by taking any word representing that element and iteratively deleting subwords abc.

✶2.9 Let B_2 be the semigroup consisting of the following five matrices:

$$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

Show that B_2 is presented by $\mathrm{Sg}\langle A \mid \sigma \cup \zeta \rangle$, where $A = \{a, b, z\}$ and

$$\sigma = \{(a^2, z), (b^2, z), (aba, a), (bab, b)\},$$
$$\zeta = \{(za, z), (az, z), (zb, z), (bz, z), (z^2, z)\}.$$

[Hint: note that the defining relations in ζ imply that z is mapped to the zero of B_2.]

2.10 Let B be the bicyclic monoid $\mathrm{Mon}\langle b, c \mid (bc, \varepsilon) \rangle$.

a) Prove that $c^\gamma b^\beta$ is idempotent if and only if $\beta = \gamma$.

b) Prove that $c^\gamma b^\beta$ is right-invertible if and only if $\gamma = 0$. [Dual reasoning will show that $c^\gamma b^\beta$ is left-invertible if and only if $\beta = 0$.]

✶2.11 Let B be the bicyclic monoid $\mathrm{Mon}\langle b, c \mid (bc, \varepsilon)\rangle$. Draw a part of the Cayley graph $\Gamma(B, \{b, c\})$ including all elements $c^\gamma b^\beta$ with $\gamma, \beta \leqslant 4$.

✶2.12 Let S be a semigroup and let $e, x, y \in S$ be such that $ex = xe = x$, $ey = ye = y$, $xy = e$, and $yx \neq e$.

a) Prove that all powers of x and all powers of y are distinct. (That is, x and y are not periodic elements.)

b) Prove that if $x^k = y^\ell$ for some $k, \ell \in \mathbb{N} \cup \{0\}$, then $k = \ell = 0$.

c) Prove that if $y^k x^\ell = e$ for some $k, \ell \in \mathbb{N} \cup \{0\}$, then $k = \ell = 0$.

d) Prove that if $y^k x^\ell = y^m x^n$ for some $k, \ell, m, n \in \mathbb{N} \cup \{0\}$, then $k = m$ and $\ell = n$.

e) Deduce that the subsemigroup $\langle x, y\rangle$ of S is isomorphic to the bicyclic monoid.

✶2.13 Let B be the bicyclic monoid and $\varphi : B \to S$ a surjective homomorphism. Prove that S is either isomorphic to the bicyclic monoid or a group.

NOTES

The section on properties of free semigroups and monoids is largely based on Howie, *Fundamentals of Semigroup Theory*, ch. 7. ◆ The discussion of semigroup presentations is partly based on Ruškuc, 'Semigroup Presentations', chs 1 & 3. ◆ For further reading on free semigroups and monoids see Harju, 'Lecture Notes on Semigroups', § 4.1–2 and Howie, *Fundamentals of Semigroup Theory*, § 7.2 on submonoids of free monoids and connections to coding theory. Lothaire, *Combinatorics on Words* is a broad study of words and contains a great deal of relevant material. For further reading on semigroup presentations, Ruškuc, 'Semigroup Presentations' is an essential text, but see also Higgins, *Techniques of Semigroup Theory*, § 1.7 & ch. 5 for an introduction to using diagrams to reason about semigroup presentations. ◆ For string-rewriting and its application to semigroup theory, see Book & Otto, *String Rewriting Systems*; for rewriting more generally, see Baader & Nipkow, *Term Rewriting and All That*. ◆ Example 2.14 is derived from the criterion for group-embeddability in Malcev, 'On the immersion of an algebraic ring into a field'. ◆ Exercise 2.5 is adapted from Gallagher, 'On the Finite Generation and Presentability of Diagonal Acts…', Proof of Proposition 3.1.12.

Structure of semigroups

3

‘ structure can be considered as a complex of relations, and ultimately as multi-dimensional order. ’

— Alfred Korzybski
Science and Sanity, bk I, pt. I, ch. 2.

The aim of this chapter is to understand better the structure of semigroups. We want to divide the semigroup into sections in such a way that we can understand the semigroup in terms of those parts and their interaction. One goal is to understand the semigroup in terms of groups; then we assume that our work is done and we hand on the problem to a group theorist.

GREEN'S RELATIONS

Green's relations

The most fundamental tools in understanding a semigroup are its Green's relations. These relate elements depending on the ideals they generate, and, as we shall see, give a lot of information about the structure of a semigroup and how its elements interact. On a semigroup, there are five Green's relations: $\mathcal{H}$, $\mathcal{L}$, $\mathcal{R}$, $\mathcal{D}$, and $\mathcal{J}$. We start by defining $\mathcal{L}$, $\mathcal{R}$, and $\mathcal{J}$: for a semigroup S, define

$\mathcal{L}$, $\mathcal{R}$, and $\mathcal{J}$

$$\left.\begin{aligned} x \mathcal{L} y &\Leftrightarrow S^1 x = S^1 y, \\ x \mathcal{R} y &\Leftrightarrow x S^1 = y S^1, \\ x \mathcal{J} y &\Leftrightarrow S^1 x S^1 = S^1 y S^1. \end{aligned}\right\} \quad (3.1)$$

It is easy to see that $\mathcal{L}$, $\mathcal{R}$, and $\mathcal{J}$ are all equivalence relations. Useful characterizations of these relations, which we will use at least as often as the definitions in (3.1), are given by the following result:

PROPOSITION 3.1. *The relations $\mathcal{L}$, $\mathcal{R}$, and $\mathcal{J}$ on a semigroup S satisfy the following:*

Characterization of $\mathcal{L}$, $\mathcal{R}$, $\mathcal{J}$

$$\begin{aligned} x \mathcal{L} y &\Leftrightarrow (\exists p, q \in S^1)((px = y) \wedge (qy = x)); \\ x \mathcal{R} y &\Leftrightarrow (\exists p, q \in S^1)((xp = y) \wedge (yq = x)); \\ x \mathcal{J} y &\Leftrightarrow (\exists p, q, r, s \in S^1)((pxr = y) \wedge (qys = x)). \end{aligned}$$

Proof of 3.1. We prove the result for $\mathcal{L}$; similar reasoning applies for $\mathcal{R}$ and $\mathcal{J}$.

Suppose $x \mathrel{\mathcal{L}} y$. Then by (3.1), $S^1x = S^1y$. Since $y \in S^1y$, it follows that $y \in S^1x$ and so there exists $p \in S^1$ such that $px = y$. Similarly, there exists $q \in S^1$ such that $qy = x$.

Now suppose that there exist $p, q \in S^1$ such that $px = y$ and $qy = x$. Then $S^1x = S^1qy \subseteq S^1y$, and similarly $S^1y = S^1px \subseteq S^1x$. Hence $S^1x = S^1y$ and so $x \mathrel{\mathcal{L}} y$. 3.1

$\mathcal{L}$ and $\mathcal{R}$ commute

PROPOSITION 3.2. $\mathcal{L} \circ \mathcal{R} = \mathcal{R} \circ \mathcal{L}$.

Proof of 3.2. Let $(x, y) \in \mathcal{L} \circ \mathcal{R}$. Then there exists $z \in S$ such that $x \mathrel{\mathcal{L}} z$ and $z \mathrel{\mathcal{R}} y$. By Proposition 3.1, there exist $p, q, r, s \in S^1$ such that $px = z$, $qz = x$, $zr = y$, and $ys = z$.

Let $z' = qzr$. Then $xr = qzr = z'$ and $z's = qzrs = qys = qz = x$, so $x \mathrel{\mathcal{R}} z'$, and $qy = qzr = z'$ and $pz' = pqzr = pxr = zr = y$, so $z' \mathrel{\mathcal{L}} y$. Hence $(x, y) \in \mathcal{R} \circ \mathcal{L}$.

Thus $\mathcal{L} \circ \mathcal{R} \subseteq \mathcal{R} \circ \mathcal{L}$. Similarly $\mathcal{R} \circ \mathcal{L} \subseteq \mathcal{L} \circ \mathcal{R}$ and so $\mathcal{L} \circ \mathcal{R} = \mathcal{R} \circ \mathcal{L}$. 3.2

As a consequence of Propositions 1.31 and 3.2, we see that $\mathcal{L} \sqcup \mathcal{R} = \mathcal{L} \circ \mathcal{R}$. Recall from page 26 that the meet of two equivalence relations is their intersection, so $\mathcal{L} \sqcap \mathcal{R} = \mathcal{L} \cap \mathcal{R}$. The meet and join of $\mathcal{L}$ and $\mathcal{R}$ play an important role, so they are also counted as Green's relations and have particular notations: $\mathcal{H}$ and $\mathcal{D}$

$$\mathcal{H} = \mathcal{L} \sqcap \mathcal{R} = \mathcal{L} \circ \mathcal{R},$$
$$\mathcal{D} = \mathcal{L} \sqcup \mathcal{R} = \mathcal{L} \cap \mathcal{R}.$$

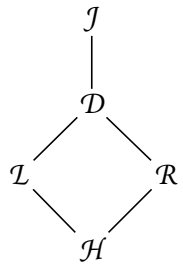


FIGURE 3.1
Hasse diagram of Green's relations in a general semigroup

From either (3.1) or Proposition 3.1, one sees that $\mathcal{L} \subseteq \mathcal{J}$ and $\mathcal{R} \subseteq \mathcal{J}$. So $\mathcal{J}$ is an upper bound for $\{\mathcal{L}, \mathcal{R}\}$ and so $\mathcal{D} = \mathcal{L} \sqcup \mathcal{R} \subseteq \mathcal{J}$. Furthermore, it is immediate that $\mathcal{H} \subseteq \mathcal{L}$ and $\mathcal{H} \subseteq \mathcal{R}$. In fact, all of these inclusions are in general strict by Exercises 3.5, 3.6, and 3.7, or by Exercise 3.11; see Figure 3.1. However, in some special classes of semigroups we do have equality of some of the relations.

For instance, let G be a group. Then in G, all of Green's relations are equal to the universal relation $G \times G$. That is, all elements of G are $\mathcal{H}$-, $\mathcal{L}$-, $\mathcal{R}$-, $\mathcal{D}$-, and $\mathcal{J}$-related.

$\mathcal{D} = \mathcal{J}$ for periodic semigroups

PROPOSITION 3.3. *In a periodic semigroup, the Green's relations $\mathcal{D}$ and $\mathcal{J}$ coincide.*

Proof of 3.3. Suppose S is periodic. We already know $\mathcal{D} \subseteq \mathcal{J}$, so we have to prove the opposite inclusion.

Let $x \mathrel{\mathcal{J}} y$. Then there exist $p, q, r, s \in S^1$ such that $pxr = y$ and $qys = x$. So $x = qpxrs$ and so $x = (qp)^n x (rs)^n$ for all $n \in \mathbb{N}$, and

similarly $y = (pq)^n y(sr)^n$ for all $n \in \mathbb{N}$. Since S is periodic, there exist $k, \ell \in \mathbb{N}$ such that $(qp)^k$ and $(sr)^\ell$ are idempotent. Let $z = px$. Then

$$\begin{aligned} x &= (qp)^k x(rs)^k = (qp)^{2k} x(rs)^k \\ &= (qp)^k((qp)^k x(rs)^k) = (qp)^k x = ((qp)^{k-1}q)z. \end{aligned}$$

Hence $x \mathcal{L} z$. Furthermore, $zr = pxr = y$ and

$$\begin{aligned} z = px &= p(qp)^{\ell+1} x(rs)^{\ell+1} = (pq)^{\ell+1} pxr(sr)^\ell s \\ &= (pq)^{\ell+1} pxr(sr)^{2\ell} s = (pq)^{\ell+1} y(sr)^{2\ell} s \\ &= (pq)^{\ell+1} y(sr)^{\ell+1}(sr)^{\ell-1} s = y((sr)^{\ell-1} s). \end{aligned}$$

Hence $z \mathcal{R} y$.

Therefore $x \mathcal{D} y$. Thus $\mathcal{J} \subseteq \mathcal{D}$ and so $\mathcal{D} = \mathcal{J}$. 3.3

PROPOSITION 3.4. a) *The relation $\mathcal{L}$ is a right congruence.*
b) *The relation $\mathcal{R}$ is a left congruence.*

Proof of 3.4. For any $x, y, z \in S$,

$$x \mathcal{L} y \Rightarrow S^1 x = S^1 y \Rightarrow S^1 xz = S^1 yz \Rightarrow xz \mathcal{L} yz,$$

and so $\mathcal{L}$ is a right congruence. Dual reasoning shows that $\mathcal{R}$ is a left congruence. 3.4

In general, $\mathcal{L}$ is not a left congruence and $\mathcal{R}$ is not a right congruence; see Exercise 3.4.

For $a \in S$, denote by H_a, L_a, R_a, D_a, and J_a the $\mathcal{H}$-, $\mathcal{L}$-, $\mathcal{R}$-, $\mathcal{D}$, and $\mathcal{J}$-classes of a, respectively. By the containment between Green's relations described above,

H_a, L_a, R_a, D_a, and J_a

$$H_a \subseteq L_a, \quad H_a \subseteq R_a, \quad L_a \subseteq D_a, \quad R_a \subseteq D_a, \quad \text{and} \quad D_a \subseteq J_a.$$

There are natural partial orders on the collection of $\mathcal{L}$-classes $S/\mathcal{L}$, the collection of $\mathcal{R}$-classes $S/\mathcal{R}$, and the collection of $\mathcal{J}$-classes $S/\mathcal{J}$ induced by inclusion order of ideals:

Partial order of $S/\mathcal{L}$, $S/\mathcal{R}$, and $S/\mathcal{J}$

$$\left.\begin{aligned} L_x \leqslant L_y &\Leftrightarrow S^1 x \subseteq S^1 y, \\ R_x \leqslant R_y &\Leftrightarrow xS^1 \subseteq yS^1, \\ J_x \leqslant J_y &\Leftrightarrow S^1 xS^1 \subseteq S^1 yS^1. \end{aligned}\right\} \quad (3.2)$$

It follows immediate from (3.2) that for all $x \in S$ and $p, q \in S^1$,

$$L_{px} \leqslant L_x, \quad R_{xq} \leqslant R_x, \quad J_{pxq} \leqslant J_x.$$

SIMPLE AND 0-SIMPLE SEMIGROUPS

Simple/0-simple

A semigroup is *simple* if it contains no proper ideals; thus S is simple if its only ideal is S itself. A semigroup S with a zero is 0-*simple* if it is not a null semigroup and its only proper ideal is $\{0\}$; thus S is 0-simple if $S^2 \neq \varnothing$ and S and $\{0\}$ are the only ideals of S.



The notion of a simple semigroup is not a generalization of a 'simple group', in the sense of a group that contains no proper non-trivial normal subgroups. Groups never contains proper ideals, so groups are always simple semigroups.

Minimal/0-minimal ideal

Let S be a semigroup and let I be an ideal (respectively, left ideal, right ideal) of S. Then I is *minimal* if there is no ideal (respectively, left ideal, right ideal) J of S that is strictly contained in I. Suppose now that S contains a zero. Then I is 0-*minimal* if $I \neq \{0\}$ and there is no ideal (respectively, left ideal, right ideal) $J \neq \{0\}$ of S that is strictly contained in I.

Uniqueness of minimal ideals

PROPOSITION 3.5. *A semigroup contains at most one minimal ideal.*

Proof of 3.5. Suppose I and J are minimal ideals of a semigroup S. Then IJ is an ideal of S and $IJ \subseteq IS \subseteq I$ and $IJ \subseteq SJ \subseteq J$. Hence, by the minimality of I and J, we have $I = IJ$ and $J = IJ$ and hence $I = J$. 3.5

Kernel of a semigroup

A semigroup S might not contain a minimal ideal. For example, the ideals I_n of $(\mathbb{N}, +)$ defined in Example 1.10(a) form an infinite descending chain: $\mathbb{N} = I_1 \supseteq I_2 \supseteq I_3 \supseteq \ldots$. But Proposition 3.5 shows that if a semigroup S contains a minimal ideal, it is unique. Such a unique minimal ideal is called the *kernel* of S and is denoted $K(S)$. Notice that if S is a semigroup with zero, $K(S) = \{0\}$.

$K(S)$

$S^2 = S$ for 0-simple semigroups

LEMMA 3.6. *If a semigroup S is 0-simple, then $S^2 = S$.*

Proof of 3.6. Note that S^2 is an ideal of S, since $SS^2 \subseteq S^2$ and $S^2S \subseteq S^2$. Now, $S^2 \neq \{0\}$ since S is not null (by the definition of 0-simple). Hence $S^2 = S$ since S is 0-simple. 3.6

LEMMA 3.7. *A semigroup S is 0-simple if and only if $SxS = S$ for all $x \in S \setminus \{0\}$.*

Proof of 3.7. Suppose S is 0-simple. Then $S^2 = S$ by Lemma 3.6 and so $S^3 = S^2S = SS = S$.

For any $x \in S$, the principal ideal SxS is either $\{0\}$ or S since S is 0-simple. Let $T = \{x \in S : SxS = \{0\}\}$. It is easy to prove that T is an ideal of S. Since S is 0-simple, it follows that $T = S$ or $T = \{0\}$. Suppose that $T = S$. Then $SxS = \{0\}$ for all $x \in S$, which implies $S^3 = \{0\}$, which is a contradiction since $S^3 = S$ by the previous paragraph. Hence $T = \{0\}$, and so $SxS = S$ for all $x \in S \setminus \{0\}$.

For the converse, suppose $SxS = S$ for all $x \in S \setminus \{0\}$. Note first that S cannot be null. Let I be some ideal of S. Suppose $I \neq \{0\}$. Then there exists some $y \in I \setminus \{0\}$, and $SyS = S$. Hence $S = SyS \subseteq I \subseteq S$ and so $I = S$. So for any ideal I of S, either $I = \{0\}$ or $I = S$, and so S is 0-simple. 3.7

PROPOSITION 3.8. a) *A 0-minimal ideal of a semigroup with a zero is either null or 0-simple.* 0-minimal ideals are 0-simple or null

b) *A minimal ideal of a semigroup is simple.*

Proof of 3.8. a) Let S be a semigroup with a zero, and let I be a 0-minimal ideal of S. Suppose I is not null. Then $I^2 \neq \{0\}$. Hence, since $I^2 \subseteq I$ is an ideal of S and I is 0-minimal, we have $I^2 = I$ and so $I^3 = I$.

Let $x \in I \setminus \{0\}$. Then S^1xS^1 is an ideal of S contained in I. Since $x \in S^1xS^1$, we have $S^1xS^1 \neq \{0\}$; hence $S^1xS^1 = I$ since I is 0-minimal. Thus $I = I^3 = IS^1xS^1I \subseteq IxI \subseteq I$. Therefore $IxI = I$ for all $x \in I \setminus \{0\}$ and so I is 0-simple by Lemma 3.7. So I is either null or 0-simple.

b) First, note that if S has a zero 0, its unique minimal ideal is $\{0\}$, which is simple. So suppose that I is a minimal ideal of a semigroup S that does not contain a zero. Then I^2 is an ideal of S and $I^2 \subseteq I$. So $I^2 = I$ since I is minimal. Hence $I^3 = I$.

Suppose J is an ideal of I. Let $x \in J$. Then $IxI \subseteq J$ since J is an ideal of I. Then S^1xS^1 is an ideal of S and $S^1xS^1 \subseteq I$; hence $S^1xS^1 = I$ since I is minimal. Therefore $J \subseteq I = I^3 = IS^1xS^1I \subseteq IxI \subseteq J$ and so $J = I$. So I is simple. 3.8

For any $x \in S$, recall that $J(x) = S^1xS^1$, and that the $\mathcal{J}$-class of x, denoted J_x, is the set of all elements of the semigroup that generate (as a principal ideal) $J(x)$. Let $I(x) = J(x) \setminus J_x$. Notice that $I(x) = \{ y \in S : J_y < J_x \}$. $I(S)$

LEMMA 3.9. *Let S be a semigroup and $x \in S$. Then $I(x)$ is either empty or an ideal of S.*

Proof of 3.9. Suppose $I(x) \neq \varnothing$. Let $y \in I(x)$ and $z \in S$. Then $yz \in J(x)$ since $J(x)$ is an ideal. But $J(yz) \subseteq J(y) \subsetneq J(x)$ (since $J(y) = J(x)$ would imply $y \in J_x$). Hence $yz \in I(x)$. Similarly $zy \in I(x)$. Hence $I(x)$ is an ideal. 3.9

The factor semigroups $J(x)/I(x)$ (where x is such that $I(x) \neq 0$) and the kernel $K(S)$ are called the *principal factors* of S. Principal factors

PROPOSITION 3.10. *Let S be a semigroup. If the kernel $K(S)$ exists, it is simple. All other principal factors of S are either null or 0-simple.* Principal factors are null or 0-simple

Proof of 3.10. By Proposition 3.8(b), if $K(S)$ exists, it is simple.

The principal factor $J(x)/I(x)$ is a 0-minimal ideal of $S/I(x)$ and so is 0-simple by Proposition 3.8(a). 3.10

Principal series

A *principal series* of a semigroup S is a finite chain of ideals

$$K(S) = S_1 \subsetneq S_2 \subsetneq \ldots \subsetneq S_n = S \tag{3.3}$$

that is maximal in the sense that there is no ideal I such that $S_i \subsetneq I \subsetneq S_{i+1}$.



Not all semigroups admit principal series. Indeed, even if a semigroup has a kernel, it may not admit a principal series: for example, let S be the semigroup $(\mathbb{N}, +)$. Then S^0 has a minimal ideal $\{0\}$ but no principal series.

We now have an analogy for semigroups of the Jordan–Hölder theorem for groups, which states that any composition series for a group contains the the same composition factors in some order.

'Jordan–Hölder theorem' for semigroups

THEOREM 3.11. *Let S be a semigroup admitting a principal series (3.3). Then the factors S_{i+1}/S_i are, in some order, isomorphic to the principal factors of S.*

Proof of 3.11. [Not especially difficult, but technical and omitted.] 3.11

$\mathcal{D}$-CLASS STRUCTURE

Since $\mathcal{L} \subseteq \mathcal{D}$ and $\mathcal{R} \subseteq \mathcal{D}$, every $\mathcal{D}$-class must be both a union of $\mathcal{L}$-classes and a union of $\mathcal{R}$-classes. On other hand, suppose that an $\mathcal{L}$-class L_x and a $\mathcal{R}$-class R_y intersect. Then there is some element $z \in L_x \cap R_y$. So $x \mathcal{L} z \mathcal{R} y$ and so $x \mathcal{D} y$. Hence L_x and R_y are both contained within the same $\mathcal{D}$-class. Therefore an $\mathcal{L}$-class and an $\mathcal{R}$-class intersect if and only if they are contained within the same $\mathcal{D}$-class.

Thus we can visualize a $\mathcal{D}$-class in the following useful way: Imagine the elements of this $\mathcal{D}$-class arranged in a rectangular pattern. This pattern is divided into a grid of cells. Each column of cells is an $\mathcal{L}$-class; each row is an $\mathcal{R}$-class, and every cell is the $\mathcal{H}$-class that is the intersection of the $\mathcal{L}$- and $\mathcal{R}$-class forming the column and row that contain that cell. This visualization is called an *egg-box diagram*; see Figure 3.2. A useful mnemonic for remembering the arrangement of an egg-box diagram is: $\mathcal{R}$-classes are $\mathcal{R}$ows and $\mathcal{L}$-classes are co$\mathcal{L}$umns. For a concrete example of an egg-box diagram, see Figure 3.7 on page 69, which is drawn using the result in Exercise 3.3.

Green's lemma

GREEN'S LEMMA 3.12. a) *Let $x, y \in S$ be such that $x \mathcal{L} y$ and let $p, q \in S^1$ be such that $px = y$ and $qy = x$. Then the 'left multiplication' maps $\lambda_p|_{R_x}$ and $\lambda_q|_{R_y}$ (where $t\lambda_z = zt$) are mutually inverse bijections between R_x and R_y. Furthermore, both of these maps preserve $\mathcal{L}$-classes, in the sense that $t\lambda_p|_{R_x} \mathcal{L} t$ and $t\lambda_q|_{R_y} \mathcal{L} t$, and so $\lambda_p|_{H_x}$ and $\lambda_q|_{H_y}$ are mutually inverse bijections between H_x and H_y.* (*See Figure 3.3.*)

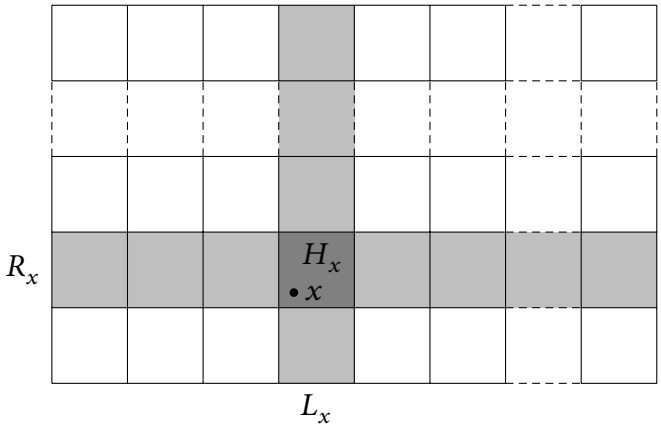


FIGURE 3.2
An egg-box diagram for the $\mathcal{D}$-class D_x. The $\mathcal{R}$-class R_x and the $\mathcal{L}$-class L_x are represented by the row and column that intersect in the box representing the $\mathcal{H}$-class H_x, which contains the element x.

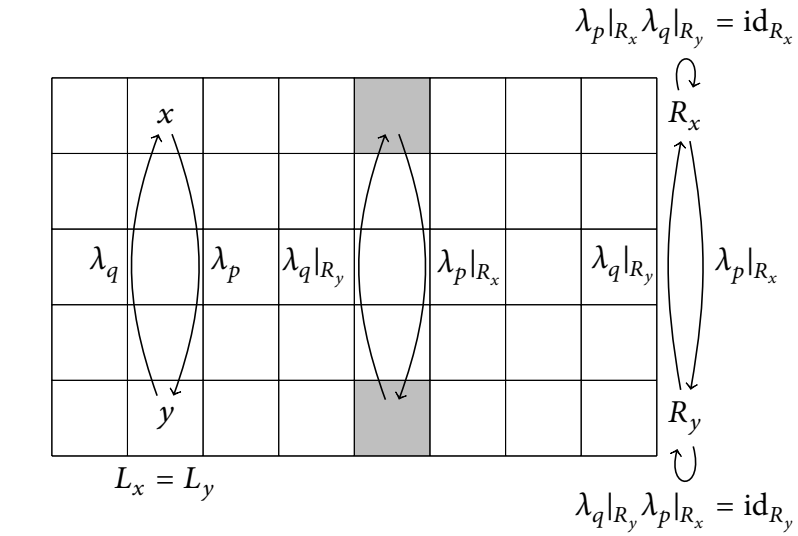


FIGURE 3.3
Green's lemma: if p and q respectively left-multiply x to give y and y to give x, then the left multiplication maps λ_p and λ_q restrict to mutually inverse bijections between the $\mathcal{R}$-classes R_x and R_y (the rows containing x and y), and both of these restricted maps preserve $\mathcal{L}$-classes (columns).

b) *Let $x, y \in S$ be such that $x \mathcal{R} y$ and let $p, q \in S^1$ be such that $xp = y$ and $yq = x$. Then the 'right multiplication' maps $\rho_p|_{L_x}$ and $\rho_q|_{L_y}$ (where $t\rho_z = tz$) are mutually inverse bijections between L_x and L_y, and both of these maps preserve $\mathcal{R}$-classes.*

Proof of 3.12. We prove only part a); the other part is proved by a dual argument.

First, notice that

$$\begin{aligned} z \in R_x &\Rightarrow z \mathcal{R} x \\ &\Rightarrow z\lambda_p|_{R_x} = pz \mathcal{R} px = y \qquad \text{[since } \mathcal{R} \text{ is a left congruence]} \\ &\Rightarrow z\lambda_p|_{R_x} \in R_y. \end{aligned}$$

So $\lambda_p|_{R_x}$ maps R_x to R_y and similarly $\lambda_q|_{R_y}$ maps R_y to R_x.

Second, suppose $z \in R_x$. Then there exists $r \in S^1$ such that $xr = z$. Then $z\lambda_p|_{R_x}\lambda_q|_{R_y} = (xr)\lambda_p|_{R_x}\lambda_q|_{R_y} = qpxr = qyr = xr = z$. Hence

$\lambda_p|_{R_x}\lambda_q|_{R_y} = \mathrm{id}_{R_x}$. Similarly $\lambda_q|_{R_y}\lambda_p|_{R_x} = \mathrm{id}_{R_y}$. So $\lambda_p|_{R_x}$ and $\lambda_q|_{R_y}$ are mutually inverse bijections.

Finally, if $z = t\lambda_p|_{R_x}$, then $z = pt$ and $t = z(\lambda_p|_{R_x})^{-1} = z\lambda_q|_{R_y} = qz$ and so $z \mathcal{L} t$. Hence $\lambda_p|_{R_x}$ preserves $\mathcal{L}$-classes. 3.12

$\mathcal{H}$-classes in the same $\mathcal{D}$-class have the same cardinality

PROPOSITION 3.13. *Let $x, y \in S$ be such that $x \mathcal{D} y$. Then $|H_x| = |H_y|$.*

Proof of 3.13. Assume $x \mathcal{D} y$. So there exists z such that $x \mathcal{L} z$ and $z \mathcal{R} y$. Let $p, q, r, s \in S^1$ be such that $px = z$, $qz = x$, $zr = y$, and $ys = z$. By Lemma 3.12, $\lambda_p|_{H_x} : H_x \to H_z$ is a bijection, and $\rho_r|_{H_z} : H_z \to H_y$ is a bijection. So $\lambda_p|_{H_x}\rho_r|_{H_z} : H_x \to H_y$ is a bijection, and hence $|H_x| = |H_y|$. 3.13

Two types of $\mathcal{H}$-class

PROPOSITION 3.14. *Let H be an $\mathcal{H}$-class of S. Then either:*

a) $H^2 \cap H = \varnothing$, *or*

b) *the following equivalent statements hold:*

i) $H^2 \cap H \neq \varnothing$;

ii) H *contains an idempotent;*

iii) $H^2 = H$;

iv) H *is a subsemigroup of S;*

v) H *is a subgroup of S.*

Proof of 3.14. If $H^2 \cap H = \varnothing$ there is nothing further to prove. So suppose that $H^2 \cap H \neq \varnothing$. Then there exist $s, t \in H$ such that $st \in H$. Then $s \mathcal{H} st$. In particular, $s \mathcal{R} st$. So by Lemma 3.12(b), $\rho_t|_H$ is a bijection from H to itself. Similarly $t \mathcal{L} st$, and thus, by Lemma 3.12(a), $\lambda_s|_H$ is a bijection from H to itself.

Now let $z \in H$. Then $sz = z\lambda_s|_H$ and $zt = z\rho_t|_H$ are both in H. Again by Lemma 3.12, $\rho_z|_H$ and $\lambda_z|_H$ are bijections from H to itself. Since $z \in H$ was arbitrary, it follows that $zH = Hz = H$ for all $z \in H$. Therefore H is a subgroup by Lemma 1.9.

We have shown that statement i) implies statement v). Statement v) clearly implies statements ii), iii), and iv), and each of these implies statement i). So all five statements are equivalent. 3.14

A *maximal subgroup* is a subgroup that does not lie inside any larger subgroup.

Maximal subgroup = $\mathcal{H}$-class containing an idempotent

PROPOSITION 3.15. *The maximal subgroups of S are precisely the $\mathcal{H}$-classes of S that contain idempotents.*

Proof of 3.15. Since every element of a subgroup is $\mathcal{H}$-related, it follows that any subgroup is contained within a single $\mathcal{H}$-class. So a maximal subgroup G is contained within a single $\mathcal{H}$-class H. But H therefore contains an idempotent 1_G and so is itself a subgroup by Proposition 3.14. Hence $H = G$. 3.15

COROLLARY 3.16. *An $\mathcal{H}$-class contains at most one idempotent.* 3.16

PROPOSITION 3.17. *Let $e \in S$ be idempotent. Then $ex = x$ for all $x \in R_e$ and $ye = y$ for all $y \in L_e$.*

Idempotents are 'left/right identities' for their $\mathcal{R}/\mathcal{L}$-classes

Proof of 3.17. Suppose $x \in R_e$. Then there exists $p \in S^1$ such that $ep = x$. Hence $ex = eep = ep = x$. Hence e is a left identity for R_e. Similarly e is a right identity for L_e. 3.17

PROPOSITION 3.18. *Let $x, y \in S$ with $x \mathcal{D} y$. Then $xy \in L_y \cap R_x$ if and only if $L_x \cap R_y$ contains an idempotent.* (*See Figure 3.4.*)

Products located by idempotents

Proof of 3.18. Suppose that $xy \in L_y \cap R_x$. In particular $xy \mathcal{R} x$. Hence there exists $q \in S^1$ such that $xyq = x$. By Lemma 3.12, $\rho_y|_{L_x} : L_x \to L_{xy}$ and $\rho_q|_{L_{xy}} : L_{xy} \to L_x$ are mutually inverse $\mathcal{R}$-class preserving bijections between L_x and L_{xy}. Since $xy \mathcal{L} y$, these maps are in fact mutually inverse $\mathcal{R}$-class preserving bijections between L_x and L_y

Hence $(yq)^2 = yqyq = y\rho_q|_{L_y}\rho_y|_{L_x}\rho_q|_{L_y} = y\rho_q|_{L_y} = yq$. Hence yq is idempotent. Furthermore, $yq = y\rho_q|_{L_y} \in L_x \cap R_y$.

Now suppose that $L_x \cap R_y$ contains an idempotent e. Then $ey = y$ by Proposition 3.17. Since $e \mathcal{R} y$, the map $\rho_y|_{L_e} : L_e \to L_y$ is an $\mathcal{R}$-class preserving bijection by Lemma 3.12. Hence $xy \in R_x \cap L_y$. 3.18

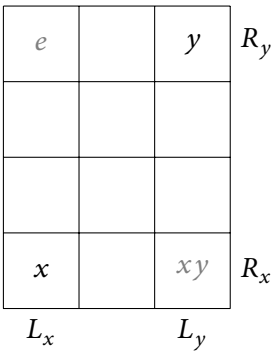


FIGURE 3.4
Products are located by idempotents: $xy \in L_y \cap R_x$ if and only if $L_x \cap R_y$ contains $e \in E(S)$.

INVERSES AND $\mathcal{D}$-CLASSES

Proposition 3.18 shows a close relationship between the product of two elements of a $\mathcal{D}$-class and idempotents in that $\mathcal{D}$-class. It is thus not surprising that idempotents and inverses in a $\mathcal{D}$-class are also connected.

PROPOSITION 3.19. *If $x \in S$ is regular, then every element of D_x is regular.*

Either every element of D_x is regular or none are

Proof of 3.19. Suppose x is regular. Then there exists $y \in S$ such that $xyx = x$. Suppose $z \mathcal{L} x$. Then there exist $p, q \in S^1$ such that $pz = x$ and $qx = z$. Hence $z = qx = qxyx = zypz$ and so z is regular. So every element of L_x is regular. A dual argument shows that if $t \in S$ is regular, every element of R_t is regular. Combining these, we see that if x is regular, every element of D_x is regular. 3.19

A $\mathcal{D}$-class is *regular* if all its elements are regular, and otherwise is *irregular*.

Regular/irregular $\mathcal{D}$-class

PROPOSITION 3.20. *In a regular $\mathcal{D}$-class, every $\mathcal{L}$-class and every $\mathcal{R}$-class contains an idempotent.*

Idempotents in a regular $\mathcal{D}$-class

Proof of 3.20. Let $x \in S$ be such that D_x is regular. In particular, x is regular and so $xyx = x$ for some $y \in S$. Now, $yx \mathcal{L} x$ and $(yx)^2 = yxyx = yx$. So yx is an idempotent in L_x. Similarly xy is an idempotent in R_x. Thus every $\mathcal{L}$-class and $\mathcal{R}$-class contains an idempotent. 3.20

Recall that $V(x)$ denotes the set of inverses of x.

PROPOSITION 3.21. *If x lies in a regular $\mathcal{D}$-class, then:*

a) *if $x' \in V(x)$, then $x \mathcal{R} xx' \mathcal{L} x'$ and $x \mathcal{L} x'x \mathcal{R} x'$ and so $x \mathcal{D} x'$;*

b) *if $z \in D_x$ is such that $L_z \cap R_x$ contains an idempotent e and $R_z \cap L_x$ contains an idempotent f, then H_z contains some $t \in V(x)$ with $xt = e$ and $tx = f$;*

c) *an $\mathcal{H}$-class contains at most one member of $V(x)$.*

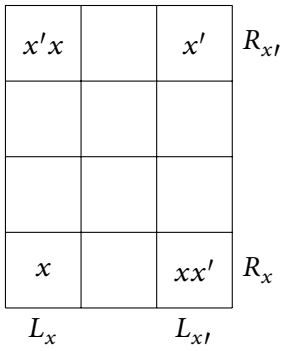


FIGURE 3.5
x and $x' \in V(x)$ in a regular $\mathcal{D}$-class

Proof of 3.21. a) Let $x' \in V(x)$. Then $xx'x = x$ and $x'xx' = x'$. Then $x \mathcal{R} xx' \mathcal{L} x'$ and so $x \mathcal{D} x'$; furthermore $x \mathcal{L} x'x \mathcal{R} x'$. (See Figure 3.5.)

b) Since $x \mathcal{R} e$, there exists $p, q \in S^1$ with $xp = e$ and $eq = x$. Let $t = fpe$. Then

$$\begin{aligned} xtx &= xfpex && \text{[by definition of } t\text{]} \\ &= xpx && \text{[since } xf = x \text{ and } ex = x \text{ by Proposition 3.17]} \\ &= ex && \text{[since } xp = e\text{]} \\ &= x && \text{[since } ex = x \text{ by Proposition 3.17]} \end{aligned}$$

and

$$\begin{aligned} txt &= fpexfpe && \text{[by choice of } t\text{]} \\ &= fpxpe && \text{[since } xf = x \text{ and } ex = x \text{ by Proposition 3.17]} \\ &= fpe^2 && \text{[since } xp = e\text{]} \\ &= fpe && \text{[since } e \text{ is idempotent]} \\ &= t. && \text{[by definition of } t\text{]} \end{aligned}$$

Hence $t \in V(x)$. Furthermore, $xt = xfpe = xpe = e^2 = e$. Finally, note that $\rho_p|_{L_x} : L_x \to L_e$ and $\rho_q|_{L_e} : L_e \to L_x$ are mutually inverse $\mathcal{R}$-class preserving bijections by Lemma 3.12(b). Hence

$$\begin{aligned} tx &= fpex && \text{[by definition of } t\text{]} \\ &= (f\rho_p|_{L_x})ex && \text{[by definition of } \rho_p|_{L_x}\text{]} \\ &= (f\rho_p|_{L_x})e^2q && \text{[since } eq = x\text{]} \\ &= (f\rho_p|_{L_x})eq && \text{[since } e \text{ is idempotent]} \\ &= (f\rho_p|_{L_x})q && \text{[by Proposition 3.17, since } f\rho_p|_{L_x} \in L_e\text{]} \\ &= f\rho_p|_{L_x}\rho_q|_{L_e} && \text{[by definition of } \rho_q|_{L_e}\text{]} \\ &= f. && \text{[since } \rho_p|_{L_x} \text{ and } \rho_q|_{L_e} \text{ are mutually inverse]} \end{aligned}$$

Now combine some of the facts we have established: from $t = fpe$ and $e = xt$, we see that $t \mathcal{L} e$; from $t = fpe$ and $f = tx$, we see that $t \mathcal{R} f$. Hence $t \in L_e \cap R_f = H_z$. (See Figure 3.6.)

c) Suppose $x', x'' \in V(x)$ and $x' \mathcal{H} x''$; we aim to show $x' = x''$. Then xx' and xx'' are idempotents lying inside $L_{x'} \cap R_x = L_{x''} \cap R_x$. Hence $xx' = xx''$ by Corollary 3.16. Similarly $x'x = x''x$. Therefore $x' = x'xx' = x'xx'' = x''xx'' = x''$. 3.21

Example 1.7 noted that every element of a rectangular band is an inverse of every element. Exercise 3.5 shows that the $\mathcal{H}$-classes of a rectangular band are the singleton sets. Thus it is possible for an element x to have an inverse in every $\mathcal{H}$-class. Exercise 3.5 also notes that a rectangular band consists of a single $\mathcal{D}$-class (which must be regular, since all elements of a rectangular band are idempotent), so all these inverses of x are $\mathcal{D}$-related to x, which fits with Proposition 3.21(a).

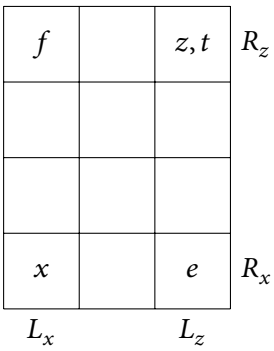


FIGURE 3.6
Inverse t corresponding to idempotents e and f in a regular $\mathcal{D}$-class

COROLLARY 3.22. *Let $e, f \in S$ be idempotents. Then $e \mathcal{D} f$ if and only if there exist $x \in S$ and $x' \in V(x)$ such that $xx' = e$ and $x'x = f$.*

Proof of 3.22. Suppose $e \mathcal{D} f$. Then $D_e = D_f$ is a regular $\mathcal{D}$-class since it contains the regular elements e and f. Choose $x \in R_e \cap L_f$ and $z \in L_e \cap R_f$. Then by Proposition 3.21(b), H_z contains some $x' \in V(x)$ such that $xx' = e$ and $x'x = f$.

Suppose now that $x \in S$ and $x' \in V(x)$ are such that $xx' = e$ and $x'x = f$. Since $e = xx'$ and $ex = xx'x = x$, it follows that $x \mathcal{R} e$. A dual argument shows that $x \mathcal{L} f$. Thus $e \mathcal{R} x \mathcal{L} f$ and so $e \mathcal{D} f$. 3.22

SCHÜTZENBERGER GROUPS

Let S be a semigroup and let H be an $\mathcal{H}$-class of S. Let $\mathrm{Stab}(H) = \{x \in S^1 : Hx = H\}$. Clearly, the adjoined identity 1 lies in $\mathrm{Stab}(H)$. If $x, y \in \mathrm{Stab}(H)$, then $Hxy = Hy = H$ and so $xy \in \mathrm{Stab}(H)$; thus $\mathrm{Stab}(H)$ is a submonoid of S^1. Define a relation σ_H on $\mathrm{Stab}(H)$ by

$\mathrm{Stab}(H)$

σ_H

$$x \sigma_H y \Leftrightarrow (\forall h \in H)(hx = hy).$$

Let $x \sigma_H y$ and $z \sigma_H t$. Let $h \in H$. Then $hx = hy$ by the definition of σ_H. Since $x, y \in \mathrm{Stab}(H)$, we have $hx = hy = h' \in H$. Thus $h'z = h't$, again by the definition of σ_H, and so $h(xz) = h(yt)$. Since $h \in H$ was arbitrary, $xz \ \sigma_H \ yt$. Therefore σ_H is a congruence on $\mathrm{Stab}(H)$. Let $\Gamma(H)$ denote the factor semigroup $\mathrm{Stab}(H)/\sigma_H$.

$\Gamma(H)$

PROPOSITION 3.23. *Let H be an $\mathcal{H}$-class of a semigroup. Then $\Gamma(H)$ is a group.*

Proof of 3.23. First of all note that $\Gamma(H)$ is a monoid with identity $[1]_{\sigma_H}$, since it is a quotient of the monoid $\mathrm{Stab}(H)$.

Let $x \in \mathrm{Stab}(H)$ and let $h \in H$. Then $hx \in H$. In particular, $hx \mathrel{\mathcal{R}} h$ and so there exists $q \in S^1$ such that $hxq = h$. Hence by Lemma 3.12, $\rho_x|_H$ and $\rho_q|_H$ are mutually inverse bijections. In particular, $Hq = H\rho_q = H$, and so $q \in \mathrm{Stab}(H)$.

Thus for any $h' \in H$, we have

$$h'xq = h'\rho_x|_H\rho_q|_H = h' = h'1,$$
$$h'qx = h'\rho_q|_H\rho_x|_H = h' = h'1.$$

Hence $xq \mathrel{\sigma_H} 1$ and $qx \mathrel{\sigma_H} 1$, and so $[x]_{\sigma_H}[q]_{\sigma_H} = [1]_{\sigma_H}$ and $[q]_{\sigma_H}[x]_{\sigma_H} = [1]_{\sigma_H}$. Since $x \in \mathrm{Stab}(H)$ was arbitrary, this proves that $\Gamma(H)$ is a group. 3.23

Schützenberger group

The group $\Gamma(H)$ is called the *Schützenberger group* of H. This notion associates a group to *every* $\mathcal{H}$-class, not just those for which $H^2 \cap H \neq \varnothing$ (see Proposition 3.14). We will see that when H is a group $\mathcal{H}$-class, $\Gamma(H)$ is actually isomorphic to H.

$\Gamma(H)$ acts regularly on H

PROPOSITION 3.24. *Let H be an $\mathcal{H}$-class of a semigroup. Then the Schützenberger group $\Gamma(H)$ acts regularly on H via $h \cdot [x]_{\sigma_H} = hx$.*

Proof of 3.24. First of all, note that the action $h \cdot [x]_{\sigma_H} = hx$ is well-defined, since if $[x]_{\sigma_H} = [y]_{\sigma_H}$, then $x \mathrel{\sigma_H} y$ and so $hx = hy$ by the definition of σ_H.

Let $h, h' \in H$. Since in particular $h \mathrel{\mathcal{R}} h'$, there exists $p \in S^1$ such that $hp = h'$. By Lemma 3.12, $\rho_p|_H$ is a bijection from H to itself, and so $p \in \mathrm{Stab}(H)$, and hence $[p]_{\sigma_H} \in \Gamma(H)$. Furthermore, $h \cdot [p]_{\sigma_H} = hp = h'$. So $\Gamma(H)$ acts transitively on H.

To show that $\Gamma(H)$ acts freely on H, we have to show that $[p]_{\sigma_H}$ is the unique element that acts on h to give h'. So suppose $h \cdot [y]_{\sigma_H} = h'$. Let $g \in H$. Since $g \mathrel{\mathcal{L}} h$, there exists $q \in S^1$ such that $qh = g$. Then

$$gy = qhy = qh \cdot [y]_{\sigma_H} = qh' = qh \cdot [p]_{\sigma_H} = qhp = gp.$$

Since this holds for all $g \in H$, it follows that $y \mathrel{\sigma_H} p$ and so $[y]_{\sigma_H} = [p]_{\sigma_H}$. Hence $\Gamma(H)$ acts freely on H.

Thus the action of $\Gamma(H)$ on H is regular. 3.24

An $\mathcal{H}$-class and its Schützenberger group have the same size

COROLLARY 3.25. *Let H be an $\mathcal{H}$-class of a semigroup. Then $|\Gamma(H)| = |H|$.*

Proof of 3.25. Since $\Gamma(H)$ acts regularly on H, there is a one-to-one correspondence between the elements of H and the elements of $\Gamma(H)$ and so $|H| = |\Gamma(H)|$. 3.25

Left Schützenberger group

Strictly speaking, $\Gamma(H)$ is the *right* Schützenberger group of H, because the definitions of $\mathrm{Stab}(H)$ and σ_H are in terms of right multiplication of elements of H. This seems arbitrary, because we could make similar definitions using left multiplication:

$$\begin{aligned}\mathrm{Stab}'(H) &= \{\, x \in S^1 : xH = H \,\};\\ x\ \sigma'_H\ y &\Leftrightarrow (\forall h \in H)(xh = yh);\\ \Gamma'(H) &= \mathrm{Stab}'(H)/\sigma'_H.\end{aligned}$$

Clearly, reasoning dual to the proofs of Propositions 3.23 and 3.24 shows that $\Gamma'(H)$ is a group that acts regularly on H *on the left* via $[x]_{\sigma'_H} \cdot h = xh$. The group $\Gamma'(H)$ is called the *left Schützenberger group* of H.

Right and left Schützenberger groups are isomorphic

PROPOSITION 3.26. $\Gamma(H) \simeq \Gamma'(H)$.

Proof of 3.26. Fix some $h \in H$. Define a map $\varphi : \Gamma(H) \to \Gamma'(H)$ as follows. For any $s \in \Gamma(H)$, since $\Gamma'(H)$ acts regularly on H, there is a unique $s' \in \Gamma'(H)$ such that $h \cdot s = s' \cdot h$. Define $s\varphi$ to be this s'. Similarly, since $\Gamma(H)$ acts regularly on H, we can define a map $\psi : \Gamma'(H) \to \Gamma(H)$ by letting $t\psi$ be the unique element of $\Gamma(H)$ such that $t \cdot h = h \cdot (t\psi)$. Clearly φ and ψ are mutually inverse and thus are bijections.

Let $[x]_{\sigma_H} \in \Gamma(H)$ and $[y]_{\sigma'_H} \in \Gamma'(H)$. Let $g \in H$. Then

$$\left.\begin{aligned}[y]_{\sigma'_H} \cdot (g \cdot [x]_{\sigma_H}) &= [y]_{\sigma'_H} \cdot (gx)\\ &= ygx\\ &= (yg) \cdot [x]_{\sigma_H}\\ &= ([y]_{\sigma'_H} \cdot g) \cdot [x]_{\sigma_H}.\end{aligned}\right\} \quad (3.4)$$

Let $s, t \in \Gamma(H)$. Then

$$\begin{aligned}(s\varphi)(t\varphi) \cdot h &= (s\varphi) \cdot (h \cdot t) && \text{[by definition of } \varphi\text{]}\\ &= \big((s\varphi) \cdot h\big) \cdot t && \text{[by (3.4)]}\\ &= (h \cdot s) \cdot t && \text{[by definition of } \varphi\text{]}\\ &= h \cdot (st) && \text{[by definition of an action; see (1.15)]}\\ &= \big((st)\varphi\big) \cdot h. && \text{[by definition of } \varphi\text{]}\end{aligned}$$

Since $\Gamma'(H)$ acts regularly on H, it follows that $(s\varphi)(t\varphi) = (st)\varphi$. Therefore φ is an isomorphism. 3.26

Schützenberger groups are the same throughout a $\mathcal{D}$-class

PROPOSITION 3.27. *Let S be a semigroup and let $x, y \in S$. If $x\ \mathcal{D}\ y$, then $\Gamma(H_x) \simeq \Gamma(H_y)$.*

Proof of 3.27. Suppose first that $x\ \mathcal{L}\ y$. Then there exist $p, q \in S^1$ such that $px = y$ and $qy = x$. So by Lemma 3.12, $\lambda_p|_{H_x} : H_x \to H_y$ and $\lambda_q|_{H_y} : H_y \to H_x$ are mutually inverse bijections. Hence $pH_x = H_y$ and $qH_y =$

H_x. Suppose that $z \in \mathrm{Stab}(H_x)$. Then $H_y z = pH_x z = pH_x = H_y$ and so $z \in \mathrm{Stab}(H_y)$. Thus $\mathrm{Stab}(H_x) \subseteq \mathrm{Stab}(H_y)$ and similarly $\mathrm{Stab}(H_y) \subseteq \mathrm{Stab}(H_x)$. So $\mathrm{Stab}(H_x) = \mathrm{Stab}(H_y)$.

Now let $z, t \in \mathrm{Stab}(H_x)$. Suppose $z\ \sigma_{H_x}\ t$. Then $xz = xt$. Let $y' \in H_y$. Since $x\ \mathcal{L}\ y'$, there exists $p' \in S^1$ such that $y' = p'x$ and so $y'z = p'xz = p'xt = y't$. Since $y' \in H_y$ was arbitrary, $z\ \sigma_{H_y}\ t$. Hence $\sigma_{H_x} \subseteq \sigma_{H_y}$. Similarly $\sigma_{H_y} \subseteq \sigma_{H_x}$ and so $\sigma_{H_x} = \sigma_{H_y}$. Therefore the Schützenberger groups $\Gamma(H_x) = \mathrm{Stab}(H_x)/\sigma_{H_x}$ and $\Gamma(H_y) = \mathrm{Stab}(H_y)/\sigma_{H_y}$ are isomorphic.

On the other hand, if $x\ \mathcal{R}\ y$, dual reasoning shows that the left Schützenberger groups $\Gamma'(H_x)$ and $\Gamma'(H_y)$ are isomorphic. The result follows from Proposition 3.26. 3.27

Notice that from Corollary 3.25 and Proposition 3.27 we immediately recover the result that if $x, y \in S$ are such that $x\ \mathcal{D}\ y$, then $|H_x| = |H_y|$ (Proposition 3.13).

PROPOSITION 3.28. *Let S be a semigroup and let H be an $\mathcal{H}$-class of S. If H is a subgroup of S, then $\Gamma(H) \simeq H$.*

Proof of 3.28. Suppose H is a group. Then $H \subseteq \mathrm{Stab}(H)$. The restriction of the natural map $\sigma_H^\natural|_H : H \to \mathrm{Stab}(H)/\sigma_H$, which maps h to $[h]_{\sigma_H}$, is a homomorphism.

Let $s \in \Gamma(H)$. Let $h = 1_H \cdot s$. Then since $1_H \cdot [h]_{\sigma_H} = h$ and $\Gamma(H)$ acts freely on H, we have $s = [h]_{\sigma_H}$. Hence $\sigma_H^\natural|_H$ is surjective.

Let $g, h \in H$ with $g\sigma_H^\natural|_H = h\sigma_H^\natural|_H$. Then $[g]_{\sigma_H} = [h]_{\sigma_H}$ and so $g = 1_H \cdot [g]_{\sigma_H} = 1_H \cdot [h]_{\sigma_H} = h$. Hence $\sigma_H^\natural|_H$ is injective.

So $\sigma_H^\natural|_H$ is an isomorphism from H to $\Gamma(H)$. Hence $\Gamma(H) \simeq H$. 3.28

Propositions 3.27 and 3.28 have the following consequence:

COROLLARY 3.29. *If H and H' are $\mathcal{H}$-classes that are subgroups within the same $\mathcal{D}$-class, then $H \simeq H'$.* 3.29

EXERCISES

[See pages 215–219 for the solutions.]

∗3.1 Prove that any two elements of a subgroup of a semigroup are $\mathcal{H}$-related.

3.2 Prove that in a free monoid A^*, we have $\mathcal{H} = \mathcal{L} = \mathcal{R} = \mathcal{D} = \mathcal{J} = \mathrm{id}_{A^*}$.

∗3.3 Let X be a set and let $\sigma, \tau \in \mathcal{T}_X$. Prove the following:

a) $\sigma\ \mathcal{L}\ \tau \Leftrightarrow \mathrm{im}\,\sigma = \mathrm{im}\,\tau$;

b) $\sigma\ \mathcal{R}\ \tau \Leftrightarrow \ker\sigma = \ker\tau$;

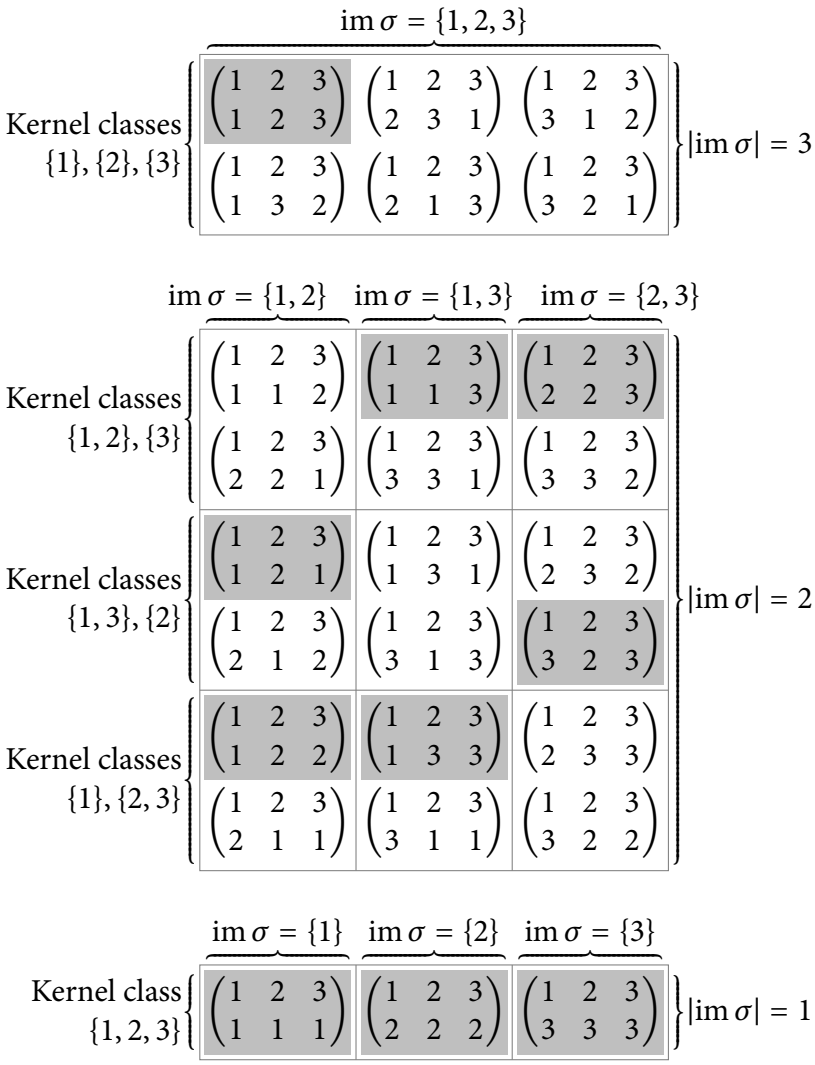


FIGURE 3.7
The egg-box diagrams of the three $\mathcal{D}$-classes of $\mathcal{T}_{\{1,2,3\}}$. Idempotents are shaded.

c) $\sigma \mathcal{D} \tau \Leftrightarrow \sigma \mathcal{J} \tau \Leftrightarrow |\operatorname{im}\sigma| = |\operatorname{im}\tau|$.

Note that the eggbox diagrams for the $\mathcal{D}$-classes of $\mathcal{T}_{\{1,2,3\}}$ are as shown in Figure 3.7.

*3.4 Give examples to show that $\mathcal{L}$ is not in general a left congruence and $\mathcal{R}$ is not in general a right congruence. [Hint: Use Exercise 3.3.]

*3.5 Let $B = L \times R$ be a rectangular band. Prove that the $\mathcal{R}$-classes of B are the sets $\{\ell\} \times R$ where $\ell \in L$, that the $\mathcal{L}$-classes of B are the sets $L \times \{r\}$ where $r \in R$, that B consists of a single $\mathcal{D}$-class, and that $\mathcal{H}$ is the identity relation.

*3.6 Prove that if S is a cancellative semigroup and does not contain an identity element, then $\mathcal{H} = \mathcal{L} = \mathcal{R} = \mathcal{D} = \mathrm{id}_S$.

*3.7 Let

$$S = \left\{ \begin{bmatrix} a & b \\ 0 & 1 \end{bmatrix} : a, b \in \mathbb{R}, a, b > 0 \right\} \subseteq M_2(\mathbb{R}).$$

Prove that S is a subsemigroup of $M_2(\mathbb{R})$. Prove that S is cancellative and has no identity, so that $\mathcal{H} = \mathcal{L} = \mathcal{R} = \mathcal{D} = \mathrm{id}_S$ by Exercise 3.6. Prove that S is simple, so that $\mathcal{J} = S \times S$.

3.8 Let $X = \{1, \ldots, n\}$ for some $n \in \mathbb{N}$. In the semigroup $\mathcal{T}_X$, prove that the $\mathcal{H}$-class containing τ is a subgroup if and only if $|\mathrm{im}\,\tau| = |\mathrm{im}(\tau^2)|$.

3.9 Recall that the bicyclic monoid B is presented by $\mathrm{Mon}\langle b, c \mid (bc, \varepsilon)\rangle$ and that every element of B has a unique representative of the form $c^\gamma b^\beta$. Prove that the $\mathcal{R}$-classes of B are sets $\{c^\gamma b^\beta : \beta \in \mathbb{N} \cup \{0\}\}$ (where $\gamma \in \mathbb{N} \cup \{0\}$ is fixed) and $\mathcal{L}$-classes of B are sets $\{c^\gamma b^\beta : \gamma \in \mathbb{N} \cup \{0\}\}$ (where $\beta \in \mathbb{N} \cup \{0\}$ is fixed). Deduce that B has a single $\mathcal{D}$-class.

3.10 Let R be an $\mathcal{R}$-class and L an $\mathcal{L}$-class of a semigroup S and suppose $L \cap R$ contains an idempotent. Let D be the $\mathcal{D}$-class containing L and R. Prove that $LR = D$.

3.11 Let M be defined by $\mathrm{Mon}\langle A \mid \rho\rangle$, where $A = \{a, b, c\}$ and $\rho = (abc, \varepsilon)$. Prove that for any $w \in M$,

$$\begin{aligned}
w \mathcal{H}\, \varepsilon &\Leftrightarrow w =_M \varepsilon;\\
w \mathcal{L}\, \varepsilon &\Leftrightarrow w \in \mathrm{Mon}\langle bc, c\rangle;\\
w \mathcal{R}\, \varepsilon &\Leftrightarrow w \in \mathrm{Mon}\langle a, ab\rangle;\\
w \mathcal{D}\, \varepsilon &\Leftrightarrow w \in \mathrm{Mon}\langle bc, c\rangle \mathrm{Mon}\langle a, ab\rangle;\\
w \mathcal{J}\, \varepsilon &\Leftrightarrow w \in \mathrm{Mon}\langle bc, c\rangle \mathrm{Mon}\langle a, ab\rangle\\
&\qquad\qquad \cup\, \mathrm{Mon}\langle bc, c\rangle b \mathrm{Mon}\langle a, ab\rangle.
\end{aligned}$$

[Hint: Recall from Exercise 2.8 that every element of M has a unique representative in $N = A^* \setminus A^* abc A^*$, and that such a representative can be obtained by iteratively deleting subwords abc.] Note that, consequently, all Green's relations are distinct in M.

3.12 Let S be a regular semigroup containing a unique idempotent. Prove that S is a group.

3.13 Let M be a group-embeddable monoid.

a) Prove that an element of x is right- and left-invertible if and only if it is $\mathcal{R}$-related to 1_M.

b) Prove that M either has one $\mathcal{R}$-class or infinitely many $\mathcal{R}$-classes.

NOTES

The exposition of Green's relations in this chapter owes much to Clifford & Preston, *The Algebraic Theory of Semigroups*, §§ 2.1–2.4 and Howie, *Fundamentals of Semigroup Theory*, §§ 2.1–2.4. The discussions of Schützenberger groups in Clifford & Preston, *The Algebraic Theory of Semigroups*, § 2.6

and Grillet, *Semigroups*, § ii.3 use a different, but equivalent, definition. ◆ The example in Exercise 3.7 is due to Andersen, 'Ein bericht über die Struktur abstrakter Halbgruppen'. ◆ The definition of the relations $\mathcal{L}$, $\mathcal{R}$, and $\mathcal{J}$, the results on principal series, Green's lemma, and the basic structure of $\mathcal{D}$-classes are all from Green, 'On the structure of semigroups'. The interaction of inverses and $\mathcal{D}$-classes is due to Miller & Clifford, 'Regular $\mathcal{D}$-classes in semigroups'. Schützenberger groups first appear, in a rather different form, in Schützenberger, '$\overline{\mathcal{D}}$ représentation des demi-groupes'. ◆ For a proof of Theorem 3.11, see Clifford & Preston, *The Algebraic Theory of Semigroups*, § 2.6. For background reading on the Jordan–Hölder theorem for groups, see Robinson, *A Course in the Theory of Groups*, § 3.1.

Regular semigroups

4

‘It looks just a little more mathematical and regular than it is; its exactitude is obvious, but its inexactitude is hidden; its wildness lies in wait.’

— G.K. Chesterton, *Orthodoxy*, ch. vi.

Properties of groups

Groups are semigroups that have many of the properties we have encountered in previous chapters. For example, groups are cancellative, regular, simple, and all of their elements have unique inverses. In this chapter, we begin to study regular semigroups, because within the class of regular semigroups there is a very interesting hierarchy of classes of semigroups that are more or less ‘group-like’, some of which have very neat structure theorems (in the sense that there is a neat description of the structure of a semigroup in this class). Figure 4.1 outlines the relationship between these classes, and it is useful to refer back to this chart to see how new definitions and results fit into the general setting. (Note that we have not yet defined many of the terms in this figure.)

Recall two basic properties of a group G: every element $x \in G$ has a unique inverse x^{-1}; and the identity 1_G is the unique idempotent, and this idempotent commutes with every element of G. A semigroup is regular if and only if every element has an inverse (Proposition 1.6), but there is no requirement that these inverses are unique: in a rectangular band, every element is an inverse of every element (Example 1.7(e)). Furthermore, every element of a rectangular band is idempotent, but they do not commute. As we shall see, the class of semigroups in which every element has a unique inverse, which are called ‘inverse semigroups’, is very important, and this turns out to be the class of regular semigroups in which idempotents commute (Theorem 5.1). If we impose another condition and require that the idempotents are central (that is, they commute with every element), we obtain the class of Clifford semigroups, which turn out to to have a very neat characterization as ‘strong semilattices of groups’. If we restrict further, and require that there is only one idempotent, we arrive at the class of groups. This is just an example; Figure 4.1 shows other ways in which we can impose properties that groups satisfy and obtain classes of more ‘group-like’ semigroups.

However, we are going to begin by defining some of these classes of semigroups in terms of properties that the inverse operation satisfies; this helps prepare the way for the study of varieties in Chapter 8. Later, we

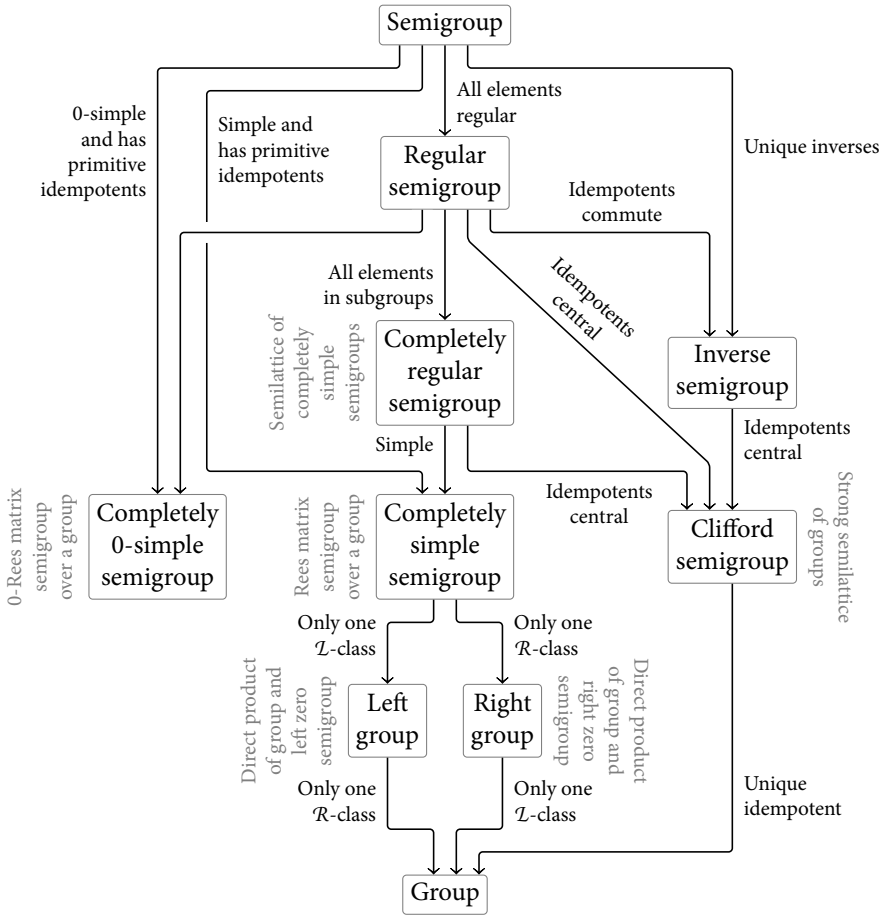


FIGURE 4.1
Chart of the classes of semigroup considered in this chapter and the following one. Labels on arrows indicate a possible extra condition (there may be others) that restricts the larger class to the smaller. Grey text summarizes the structure theorem for the adjacent class. (Some of these classes have not yet been defined.)

will show how these classes fit into the chart in Figure 4.1.

Let S be a semigroup. If S is a group, the map $x \mapsto x^{-1}$ that sends an element to its inverse is a unary operation on S that satisfies certain properties. For instance, by definition $xx^{-1} = x^{-1}x = 1_S$ for all $x \in S$. But $^{-1}$ also satisfies other properties: for all $x, y \in S$,

$$(x^{-1})^{-1} = x, \qquad (xy)^{-1} = y^{-1}x^{-1}, \qquad x^{-1}x = xx^{-1},$$
$$xx^{-1}x = x, \qquad xx^{-1}yy^{-1} = yy^{-1}xx^{-1}, \qquad xx^{-1} = yy^{-1}.$$

If we require that a semigroup with an operation $^{-1}$ satisfies only some of these properties, we may no longer have a group. Instead, we obtain different types of semigroup depending on which conditions are required.

Regular semigroup

Let S be a semigroup equipped with an operation $^{-1}$. If S satisfies the condition that for all $x \in S$,

$$xx^{-1}x = x,$$

then S is clearly regular, as defined on page 6. [Note that in a regular semigroup, an element may have many different inverses. However, we can always define an operation $^{-1}$ by choosing a particular inverse for

each element.] If S satisfies the two conditions that for all $x \in S$,

$$(x^{-1})^{-1} = x, \qquad xx^{-1}x = x, \tag{4.1}$$

then again S is regular and for any $y \in S$, we have $yy^{-1}y = y$ and $y^{-1}yy^{-1} = y^{-1}(y^{-1})^{-1}y^{-1} = y^{-1}$ and so y^{-1} is an inverse of y. If S satisfies the three conditions that for all $x \in S$,

Completely regular semigroup

$$(x^{-1})^{-1} = x, \qquad x^{-1}x = xx^{-1}, \qquad xx^{-1}x = x, \tag{4.2}$$

it is a *completely regular semigroup*. We will look at regular and completely regular semigroups in this chapter. If S satisfies the four conditions that for all $x, y \in S$,

Inverse semigroup

$$\left.\begin{aligned} (x^{-1})^{-1} &= x, & (xy)^{-1} &= y^{-1}x^{-1}, \\ xx^{-1}x &= x, & xx^{-1}yy^{-1} &= yy^{-1}xx^{-1}, \end{aligned}\right\} \tag{4.3}$$

it is an *inverse semigroup*. Finally, if S satisfies the four conditions that for all $x, y \in S$,

Clifford semigroup

$$\left.\begin{aligned} (x^{-1})^{-1} &= x, & x^{-1}x &= xx^{-1}, \\ xx^{-1}x &= x, & xx^{-1}yy^{-1} &= yy^{-1}xx^{-1}, \end{aligned}\right\} \tag{4.4}$$

it is a *Clifford semigroup*. We will look at inverse semigroups and Clifford semigroups in Chapter 5.

COMPLETELY 0-SIMPLE SEMIGROUPS

The aim of this section is to introduce the concept of a completely 0-simple semigroup and to present a classification result for such semigroups, the Rees–Suschkewitsch theorem, which was one of the most important results in the early development of semigroup theory. We study completely 0-simple semigroups for two reasons. First, we saw in Proposition 3.10 that the principal factors of a semigroup are either 0-simple or null, and completely 0-simple semigroup are an important subclass of 0-simple semigroups. Furthermore, studying completely 0-simple semigroups will lead naturally to studying completely simple semigroups, and we will see that both completely 0-simple and completely simple semigroups are regular (Lemma 4.6(b) and Proposition 4.13), and that a simple semigroup is completely simple if and only if it is completely regular (Theorem 4.16).

Recall that the set of idempotents $E(S)$ of a semigroup S admits a natural partial order given by $e \preccurlyeq f \Leftrightarrow ef = fe = e$. (See Proposition 1.19.) In a semigroup with a zero, 0 is the unique minimal idempotent; in such a semigroup, an idempotent is *primitive* if it is minimal within the set of non-zero idempotents of the semigroup. A semigroup is *completely 0-simple* if it is 0-simple and contains at least one primitive idempotent.

Primitive idempotent

Completely 0-simple semigroup

Finite 0-simple ⇒ completely 0-simple

PROPOSITION 4.1. *A finite* 0*-simple semigroup is completely* 0*-simple.*

Proof of 4.1. Let S be a finite 0-simple semigroup. Now, if there are non-zero idempotents in S, there must be a primitive idempotent in S, since otherwise there would be infinite descending chains of idempotents, and this is impossible since S is finite. So we must simply rule out the possibility that 0 is the only idempotent.

So suppose, with the aim of obtaining a contradiction, that the only idempotent in S is 0. Let $x \in S \setminus \{0\}$. Then by Lemma 3.7 there exist $p, q \in S$ with $pxq = x$. Hence $p^n x q^n = x$ for all $n \in \mathbb{N}$. Since S is finite and thus periodic, p^m is idempotent for some $m \in \mathbb{N}$. Thus $p^m = 0$ since 0 is the only idempotent in S. Therefore $x = p^m x q^m = 0xq^m = 0$, which contradicts the choice of x. So it is impossible for 0 to be the only idempotent of S. This completes the proof. 4.1

Rees matrix semigroup

We are now going to show how to construct examples of completely 0-simple semigroups. Let G be a group, let I and Λ be abstract index sets, and let P be a regular $\Lambda \times I$ matrix with entries from G^0. (Recall that a matrix is *regular* if every row and every column contains at least one non-zero entry. By a '$\Lambda \times I$ matrix' we mean simply a matrix whose rows are indexed by Λ and whose columns are indexed by I.) Let $p_{\lambda i}$ be the (λ, i)-th entry of P. Let S be the set $I \times G^0 \times \Lambda$. Define a multiplication on S by

$$(i, x, \lambda)(j, y, \mu) = (i, x p_{\lambda j} y, \mu).$$

This multiplication is associative since

$$\begin{aligned}(i, x, \lambda)\big((j, y, \mu)(k, z, \nu)\big) &= (i, x, \lambda)(j, y p_{\mu k} z, \nu)\\ &= (i, x p_{\lambda j} y p_{\mu k} z, \nu)\\ &= (i, x p_{\lambda j} y, \mu)(k, z, \nu)\\ &= \big((i, x, \lambda)(j, y, \mu)\big)(k, z, \nu),\end{aligned}$$

and so S is a semigroup. Let $T = I \times \{0\} \times \Lambda$. It is easy to see that T is an ideal of S. Clearly, $S \setminus T = I \times G \times \Lambda$. Notice that if $(i, x, \lambda), (j, y, \mu) \in S \setminus T$, then $(i, x, \lambda)(j, y, \mu) \in T$ if and only if $p_{\lambda j} = 0$.

Let $\mathcal{M}_0[G; I, \Lambda; P]$ be the Rees factor semigroup S/T. Then the semigroup $\mathcal{M}_0[G; I, \Lambda; P]$ can be viewed as the set $(S \setminus T) \cup \{0\}$: that is, $\mathcal{M}_0[G; I, \Lambda; P]$ can be viewed as the set $(I \times G \times \Lambda) \cup \{0\}$ under the multiplication

$$(i, x, \lambda)(j, y, \mu) = \begin{cases}(i, x p_{\lambda j} y, \mu) & \text{if } p_{\lambda j} \neq 0,\\ 0 & \text{if } p_{\lambda j} = 0,\end{cases}$$

$$0(i, x, \lambda) = (i, x, \lambda)0 = 00 = 0.$$

The semigroup $\mathcal{M}_0[G; I, \Lambda; P]$ is called the $I \times \Lambda$ *Rees matrix semigroup over* G^0 *with regular sandwich matrix* P.

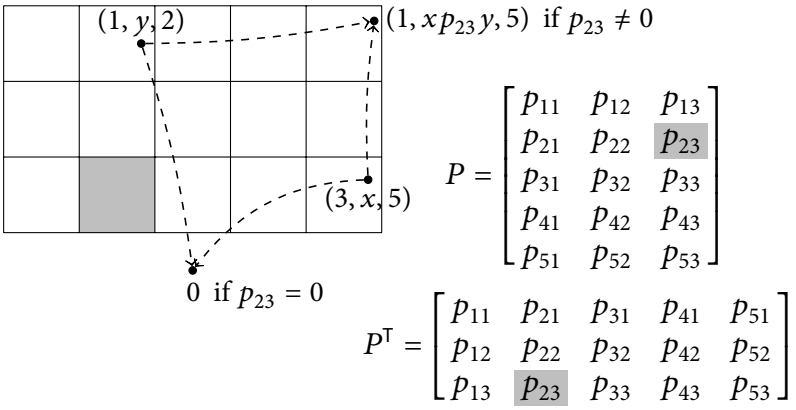


FIGURE 4.2
Multiplication in a Rees matrix semigroup $\mathcal{M}_0[G; I, \Lambda; P]$. The product $(1, y, 2)(3, x, 5)$ is either $(1, yp_{23}x, 5)$ or 0, depending on the value of p_{23}. The shape of P^T is the same as the shape of the grid, and the cells containing the multiplicands, the cell corresponding to p_{23}, and cell containing the product (if it is non-zero) form the corners of a rectangle.

Diagrammatically, we can place the non-zero elements of this Rees matrix semigroup in a rectangular pattern, divided into a grid of cells indexed by the sets I and Λ, so that the (i, λ)-th cell contains all elements of the form (i, g, λ), where $g \in G$. Figure 4.2 illustrates how the multiplication works in terms of this diagram. Compare this with Figure 1.1. This is of course reminiscent of an egg-box diagram, and we will see that it actually *is* an egg-box diagram: the columns, rows, and cells of this grid are the non-zero $\mathcal{L}$-, $\mathcal{R}$-, and $\mathcal{H}$-classes of the Rees matrix semigroup.

PROPOSITION 4.2. *For any group G, index sets I and Λ, and matrix P over G^0, the semigroup $\mathcal{M}_0[G; I, \Lambda; P]$ is completely 0-simple.*

Rees matrix ⇒ completely 0-simple

Proof of 4.2. For brevity, let $S = \mathcal{M}_0[G; I, \Lambda; P]$.

Let $(i, x, \lambda) \in S \setminus \{0\}$. Let $(j, y, \mu) \in S \setminus \{0\}$. Since P is regular, we can choose $\nu \in \Lambda$ and $k \in I$ such that $p_{\nu i} \neq 0$ and $p_{\lambda k} \neq 0$. Then

$$\begin{aligned}
&(j, 1_G, \nu)(i, x, \lambda)(k, p_{\lambda k}^{-1} x^{-1} p_{\nu i}^{-1} y, \mu) \\
&= (j, 1_G p_{\nu i} x p_{\lambda k} p_{\lambda k}^{-1} x^{-1} p_{\nu i}^{-1} y, \mu) \\
&= (j, y, \mu).
\end{aligned}$$

Hence, since $(j, y, \mu) \in S \setminus \{0\}$ was arbitrary, and since $0 = 0(i, x, \lambda)0$, we have $S \subseteq S(i, x, \lambda)S$. Since $(i, x, \lambda) \in S \setminus \{0\}$ was arbitrary, S is 0-simple by Lemma 3.7.

Now, $(i, x, \lambda) \in S \setminus \{0\}$ is an idempotent if and only if $(i, x, \lambda)(i, x, \lambda) = (i, xp_{\lambda i}x, \lambda) = (i, x, \lambda)$, which is true if and only if $p_{\lambda i} \neq 0$ and $x = p_{\lambda i}^{-1}$. Hence the idempotents in $S \setminus \{0\}$ are elements of the form $(i, p_{\lambda i}^{-1}, \lambda)$. Furthermore,

$$\begin{aligned}
&(i, p_{\lambda i}^{-1}, \lambda) \preccurlyeq (j, p_{\mu j}^{-1}, \mu) \\
\Leftrightarrow\ &(i, p_{\lambda i}^{-1}, \lambda)(j, p_{\mu j}^{-1}, \mu) = (j, p_{\mu j}^{-1}, \mu)(i, p_{\lambda i}^{-1}, \lambda) = (i, p_{\lambda i}^{-1}, \lambda) \\
\Leftrightarrow\ &(i, p_{\lambda i}^{-1} p_{\lambda j} p_{\mu j}^{-1}, \mu) = (j, p_{\mu j}^{-1} p_{\mu i} p_{\lambda i}^{-1}, \lambda) = (i, p_{\lambda i}^{-1}, \lambda) \\
\Leftrightarrow\ &(i = j) \wedge (\lambda = \mu) \\
\Leftrightarrow\ &(i, p_{\lambda i}^{-1}, \lambda) = (j, p_{\mu j}^{-1}, \mu).
\end{aligned}$$

Hence every idempotent in $S \setminus \{0\}$ is primitive. Thus S certainly contains primitive idempotents and so is completely 0-simple. 4.2

Proposition 4.2 gives a method for constructing completely 0-simple semigroups. In fact, *all* completely 0-simple semigroups arise in this way:

Completely 0-simple ⇒ Rees matrix

PROPOSITION 4.3. *Let S be a completely 0-simple semigroup. Then $S \simeq \mathcal{M}_0[G; I, \Lambda; P]$ for some group G, index sets I and Λ, and regular sandwich matrix P.*

Proof of 4.3. Let S be completely 0-simple. We have to define a Rees matrix semigroup $\mathcal{M}_0[G; I, \Lambda; P]$ and show it is isomorphic to S.

Since S is completely 0-simple, it contains a primitive idempotent. We first describe the $\mathcal{R}$-classes and $\mathcal{L}$-classes of primitive idempotents:

LEMMA 4.4. *For any primitive idempotent e of S,*

a) $R_e = eS \setminus \{0\}$,

b) $L_e = Se \setminus \{0\}$.

Proof of 4.4. We prove part a); a dual argument gives part b). Note that by definition $e \neq 0$. Every element of R_e is a right multiple of e and cannot be 0. Hence $R_e \subseteq eS \setminus \{0\}$.

Let $x \in eS \setminus \{0\}$. So $x = es$ for some $s \in S \setminus \{0\}$. Hence $ex = ees = es = x$. Since S is 0-simple, by Lemma 3.7 there exist $p, q \in S$ with $pxq = e$. Let $p' = epe$. Then $p'xq = epexq = epxq = ee = e$.

Let $f = xqp'$. Then $f^2 = xqp'xqp' = xqep' = xqeepe = xqepe = xqp' = f$. So f is idempotent. Furthermore, $ef = exqp' = xqp' = f$ and $fe = xqp'e = xqepee = xqepe = xqp' = f$. So $ef = fe = f$ and hence $f \preccurlyeq e$. Suppose that $f = 0$; then $e = e^2 = p'xqp'xq = p'fxq = 0$, which is a contradiction. Hence $f \neq 0$. But e is primitive and therefore $\preccurlyeq$-minimal among non-zero idempotents; thus $e = f = xqp'$. Since $x = es$, it follows that $x \mathcal{R} e$ and so $x \in R_e$. Hence $eS \setminus \{0\} \subseteq R_e$. 4.4

LEMMA 4.5. *For any $x \in S \setminus \{0\}$,*

a) $R_x = xS \setminus \{0\}$,

b) $L_x = Sx \setminus \{0\}$.

Proof of 4.5. We prove part a); a dual argument gives part b). As in the proof of Lemma 4.4(a), $R_x \subseteq xS \setminus \{0\}$. So let $y \in xS \setminus \{0\}$. Then $y = xs$ for some $s \in S \setminus \{0\}$. Let e be a primitive idempotent of S. Since S is 0-simple, by Lemma 3.7 there exist $p, q \in S$ with $peq = x$. So $y = peqs$. By Lemma 4.4(a), $eqs, eq \in R_e$. Since $\mathcal{R}$ is a left congruence by Propostion 3.4, $y = peqs \mathcal{R} peq = x$. So $y \in R_x$ and hence $xS \setminus \{0\} \subseteq R_x$. 4.5

We can now deduce information about the $\mathcal{D}$-class structure of S:

LEMMA 4.6. a) *The $\mathcal{D}$-classes of S are 0 and $S \setminus \{0\}$.*

b) *The semigroup S is regular.*

c) *For all* $x, y \in S \setminus \{0\}$, *if* $L_x \cap R_y$ *contains an idempotent, then* $xy \in R_x \cap L_y$; *otherwise,* $xy = 0$.

Proof of 4.6. a) Let $x, y \in S \setminus \{0\}$. Suppose $xSy = \{0\}$. Then

$$S^2 = SxSSySS \subseteq S(xSy)S = S\{0\}S = \{0\},$$

which contradicts the fact that 0-simple semigroups are (by definition) not null. Hence xSy contains some non-zero element t. Now, $t \in xS \setminus \{0\}$ and $t \in Sy \setminus \{0\}$. Hence $t \in R_x$ and $t \in L_y$ by Lemma 4.5. Thus $x \mathcal{R} t \mathcal{L} y$ and so $x \mathcal{D} y$. So the $\mathcal{D}$-classes of S must be $S \setminus \{0\}$ and $\{0\}$.

b) The primitive idempotent e lies in $S \setminus \{0\}$ and so every element of $S \setminus \{0\}$ is regular by Proposition 3.19. Since 0 is also regular, the semigroup S is regular.

c) Let $x, y \in S \setminus \{0\}$. By part b), $x \mathcal{D} y$. Suppose $L_x \cap R_y$ contains an idempotent. Then $xy \in R_x \cap L_y$ by Proposition 3.18. Suppose $L_x \cap R_y$ does not contain an idempotent. Then $xy \notin R_x \cap L_y$, and so $xy = 0$, since if $xy = 0$, then by Lemma 4.6 $xy \in xS \setminus \{0\} = R_x$ and $xy \in Sy \setminus \{0\} = L_y$, contradicting $xy \notin R_x \cap L_y$. 4.6

Let H be an $\mathcal{H}$-class of S contained in the $\mathcal{D}$-class $S \setminus \{0\}$. Let $x, y \in H$. Then either $xy = 0$ or $xy \in R_x \cap L_y = H$ by Lemma 4.6(c).

Suppose first that $xy = 0$. Let $z, t \in H$. Since $z \mathcal{L} x$ and $t \mathcal{R} y$, we have $z = px$ and $t = ys$ for some $p, s \in S^1$. Then $zt = pxys = p0r = 0$. Since $z, t \in H$ were arbitrary, $H^2 = \{0\}$.

On the other hand, suppose that $xy \in H$. Then H is a subgroup by Proposition 3.14. So we can divide the $\mathcal{H}$-classes in the $\mathcal{D}$-class $S \setminus \{0\}$ into *zero* $\mathcal{H}$-classes and *group* $\mathcal{H}$-classes.

Let I be the set of $\mathcal{R}$-classes and let Λ be the set of $\mathcal{L}$-classes in $S \setminus \{0\}$. Write the $\mathcal{R}$- and $\mathcal{L}$-classes as $R^{(i)}$ and $L^{(\lambda)}$ for $i \in I$ and $\lambda \in \Lambda$, and write $H^{(i\lambda)}$ for $R^{(i)} \cap L^{(\lambda)}$. We will treat I and Λ as abstract index sets, and these will ultimately be the index sets for the Rees matrix semigroup we are constructing.

Since $S \setminus \{0\}$ is a regular $\mathcal{D}$-class by Lemma 4.6(b), every $\mathcal{R}$-class and every $\mathcal{L}$-class contains an idempotent by Proposition 3.20 and thus contains some group $\mathcal{H}$-class. Therefore assume without loss that there is some element $1 \in I \cap \Lambda$ such that $H^{(11)}$ is a group $\mathcal{H}$-class. For brevity, write H for $H^{(11)}$.

For each $i \in I$ and $\lambda \in \Lambda$, fix arbitrary elements $r_\lambda \in H^{(1\lambda)} \subseteq R^{(1)}$ and $q_i \in H^{(i1)} \subseteq L^{(1)}$. Since 1_H is idempotent, it is a left identity for $R^{(1)}$ and a right identity for $L^{(1)}$ by Proposition 3.17. So $1_H r_\lambda = r_\lambda$ and $q_i 1_H = q_i$. Therefore, by Lemma 3.12, $\rho_{r_\lambda}|_{L^{(1)}} : L^{(1)} \to L^{(\lambda)}$ restricts to a bijection between H and $H^{(i1)}$ and $\lambda_{q_i}|_{R^{(1)}} : R^{(1)} \to R^{(1)}$ restricts to a bijection between $H^{(i1)}$ and $H^{(i\lambda)}$ for each $\lambda \in \Lambda$. Thus there is a unique expression $q_i x r_\lambda$, where $x \in H$, for every element of $H^{(i\lambda)}$.

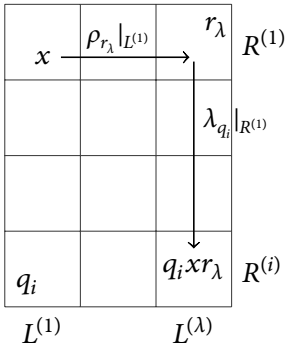


FIGURE 4.3
Choosing $r_\lambda \in H^{(1\lambda)}$ and $q_i \in H^{(i1)}$ gives bijections $\rho_{r_\lambda}|_{L^{(1)}}$ and $\lambda_{q_i}|_{R^{(1)}}$ and so a unique expression $q_i x r_\lambda$ for each element of $H^{(i\lambda)}$.

Therefore the map $\varphi : (I \times H \times \Lambda) \cup \{0\} \to S$ defined by $(i, x, \lambda)\varphi = q_i x r_\lambda$ and $0\varphi = 0$ is a bijection.

To turn $(I \times H \times \Lambda) \cup \{0\}$ into a $I \times \Lambda$ Rees matrix semigroup over H^0, it remains to define a sandwich matrix P. For each $i \in I$ and $\lambda \in \Lambda$, let $p_{\lambda i} = r_\lambda q_i$, and let P be the $\Lambda \times I$ matrix whose (λ, i)-th entry is $p_{\lambda i}$. By Lemma 4.6(c), $p_{\lambda i} = r_\lambda q_i \in R_{r_\lambda} \cap L_{q_i} = R^{(1)} \cap L^{(1)} = H$ if and only if $R_{q_i} \cap L_{r_\lambda}$ contains an idempotent and is thus a group $\mathcal{H}$-class; otherwise $p_{\lambda i} = 0$. Hence each $p_{\lambda i}$ lies in H^0. Furthermore, since every $\mathcal{R}$-class and every $\mathcal{L}$-class contains an idempotent, for every $i \in I$ there exists $\lambda \in \Lambda$ such that $R_{q_i} \cap L_{r_\lambda}$ contains an idempotent and so $p_{\lambda i} \in H$, and thus $p_{\lambda i} \neq 0$. Thus every column of P contains a non-zero entry. Similarly every row of O contains a non-zero entry. Therefore P is a regular matrix.

So φ is now a bijection from $\mathcal{M}_0[H; I, \Lambda; P]$ to S. For any elements $(i, x, \lambda), (j, y, \mu) \in \mathcal{M}_0[H; I, \Lambda; P] \setminus \{0\}$,

$$\begin{aligned}
\big((i, x, \lambda)\varphi\big)\big((j, y, \mu)\varphi\big) &= (q_i x r_\lambda)(q_j y r_\mu) \\
&= q_i x (r_\lambda q_j) y r_\mu \\
&= q_i x p_{\lambda j} y r_\mu \\
&= \begin{cases} (i, x p_{\lambda j} y, \mu)\varphi & \text{if } p_{\lambda j} \neq 0 \\ 0\varphi & \text{if } p_{\lambda j} = 0 \end{cases} \\
&= \big((i, x, \lambda)(j, y, \mu)\big)\varphi.
\end{aligned}$$

Furthermore, $\big((i, x, \lambda)\varphi\big)(0\varphi) = q_i x r_\lambda 0 = 0 = \big((i, x, \lambda)0\big)\varphi$ and similarly for other multiplications involving 0. Hence the map φ is a homomorphism and hence an isomorphism between $\mathcal{M}_0[H; I, \Lambda; P]$ and S. 4.6

Combining Propositions 4.2 and 4.3, we get the following characterization of completely 0-simple semigroups:

Rees–Suschkewitsch theorem

REES–SUSCHKEWITSCH THEOREM 4.7. *A semigroup S is completely 0-simple if and only if there exist a group G, index sets, I and Λ, and a regular $\Lambda \times I$ matrix P with entries from G^0 such that $S \simeq \mathcal{M}_0[G; I, \Lambda; P]$.* 4.7

One of the advantages of this characterization is that it gives us a neat description of the $\mathcal{H}$, $\mathcal{L}$, $\mathcal{R}$, $\mathcal{D}$, and $\mathcal{J}$-classes:

Green's relations in completely 0-simple semigroups

PROPOSITION 4.8. *Let $S \simeq \mathcal{M}_0[G; I, \Lambda, P]$ be a completely 0-simple semigroup.*

a) *In S, the relations $\mathcal{D}$ and $\mathcal{J}$ coincide, and S has two $\mathcal{D}$-classes $\{0\}$ and $S \setminus \{0\} = I \times G \times \Lambda$.*

b) *The $\mathcal{L}$-classes of S are $\{0\}$ and sets of the form $I \times G \times \{\lambda\}$.*

c) *The $\mathcal{R}$-classes of S are $\{0\}$ and sets of the form $\{i\} \times G \times \Lambda$.*

d) *The $\mathcal{H}$-classes of S are $\{0\}$ and sets of the form $\{i\} \times G \times \{\lambda\}$.*

Proof of 4.8. a) By Lemma 4.6, S has two $\mathcal{D}$-classes, $\{0\}$ and $S \setminus \{0\}$. Since S is 0-simple, it has only two ideals: S and $\{0\}$. If $x \in S \setminus \{0\}$, then $x \in S^1 x S^1$, so $S^1 x S^1 = S$. On the other hand, $S^1 0 S^1 = \{0\}$. So S and $\{0\}$ are also the $\mathcal{J}$-classes of S.

b) Since $\{0\}$ is the $\mathcal{D}$-class of 0, it is alsoe the $\mathcal{L}$-class of 0.

Let $(i, x, \lambda) \setminus \{0\}$. By Lemma 4.5(b), we have $L_{(i,x,\lambda)} = S(i, x, \lambda) \setminus \{0\}$. First, note that $S(i, x, \lambda) \setminus \{0\} \subseteq I \times G \times \{\lambda\} \setminus \{0\}$ by the definition of multiplication in $\mathcal{M}_0[G; I, \Lambda; P]$.

On the other hand, let $(j, y, \lambda) \in I \times G \times \{\lambda\} \setminus \{0\}$. Let $s = (j, yx^{-1} p_{\mu i}^{-1}, \mu)$, where μ is such that $p_{\mu i} \neq 0$; such a μ exists because P is regular. Note that $p_{\mu i} \in G$, so $p_{\mu i}^{-1}$ exists. Then

$$\begin{aligned} s(i, x, \lambda) &= (j, yx p_{\mu i}^{-1}, \mu)(i, x, \lambda) \\ &= (j, yx^{-1} p_{\mu i}^{-1} p_{\mu i} x, \lambda) \\ &= (j, y, \lambda). \end{aligned}$$

So $I \times G \times \{\lambda\} \setminus \{0\} \subseteq S(i, x, \lambda) \setminus \{0\}$.

Thus $L_{(i,x,\lambda)} = I \times G \times \{\lambda\} \setminus \{0\}$.

c) The reasoning is dual to part b).

d) First, $H_0 = L_0 \cap R_0 = \{0\}$. For $(i, x, \lambda) \in S \setminus \{0\}$, we have $H_{(i,x,\lambda)} = L_{(i,x,\lambda)} \cap R_{(i,x,\lambda)} = (I \times G \times \{\lambda\}) \cap (\{i\} \times G \times \Lambda) = \{i\} \times G \times \{\lambda\}$. 4.8

IDEALS AND COMPLETELY 0-SIMPLE SEMIGROUPS

This section characterizes the 0-simple semigroups that are also completely 0-simple. We require some definitions. A semigroup S is *group-bound* if every $x \in S$ has some power x^n lying in a subgroup of S. A semigroup *satisfies the condition* $\min_{\mathcal{L}}$ (respectively, $\min_{\mathcal{R}}$) if any subset of the partial order $S/\mathcal{L}$ (respectively, $S/\mathcal{R}$) has a minimal element.

Group-bound semigroup

$\min_{\mathcal{L}}$, $\min_{\mathcal{R}}$

THEOREM 4.9. *Let S be 0-simple. The following are equivalent:*

a) *S is completely 0-simple;*

b) *S is group-bound;*

c) *S satisfies the conditions $\min_{\mathcal{L}}$ and $\min_{\mathcal{R}}$.*

Characterization of 0-simple semigroups that are completely 0-simple

Proof of 4.9. Part 1 [a) $\Rightarrow$ b)]. Suppose S is completely 0-simple. Let $x \in S$. Then either H_x is a subgroup and $x^2 \in H_x$, or else $x^2 = 0$. In either case, x^2 lies in a subgroup. Thus S is group-bound.

Part 2 [b) $\Rightarrow$ c)]. Suppose S is group-bound. Let $x, y \in S \setminus \{0\}$ be such that $L_x \leqslant L_y$. We are going to show that $L_x = L_y$. First, notice that $x = py$ for some $p \in S^1$. Furthermore, $y = qxr$ for some

$q, r \in S$ by Lemma 3.7, since S is 0-simple. Then $y = qxr = qpyr$ and so $y = (qp)^n yr^n$ for all $n \in \mathbb{N}$. Fix n so that $g = (qp)^n$ lies in a subgroup G. Then $1_G y = 1_G gyr^n = gyr^n = y$ and so $y = g^{-1}gy = g^{-1}(qp)^n y = g^{-1}(qp)^{n-1}qpy = g^{-1}(qp)^{n-1}qx$. Hence $L_y \leqslant L_x$ and so $L_x = L_y$. Therefore $L_x \leqslant L_y \Rightarrow L_x = L_y$, and this certainly implies that any subset of $S/\mathcal{L}$ is has a minimal element. So S satisfies $\min_{\mathcal{L}}$. Similarly, $S/\mathcal{R}$ satisfies $\min_{\mathcal{R}}$.

Part 3 [c) $\Rightarrow$ a)]. Suppose S satisfies $\min_{\mathcal{L}}$ and $\min_{\mathcal{R}}$. Suppose, with the aim of obtaining a contradiction, that S does not contain a primitive idempotent. Then S contains an infinite descending chain of non-zero idempotents

$$e_1 \succ e_2 \succ e_3 \succ \dots .$$

Notice that for $e, f \in E(S)$, by the definition of the partial order $\preccurlyeq$ on $E(S)$ and the relation $\mathcal{R}$, we have

$$e \preccurlyeq f \Rightarrow e = fe \Rightarrow eS = feS \subseteq fS \Rightarrow R_e \leqslant R_f$$

and similarly $e \preccurlyeq f \Rightarrow L_e \leqslant L_f$. Hence

$$L_{e_1} \geqslant L_{e_2} \geqslant L_{e_3} \geqslant \dots \quad \text{and} \quad R_{e_1} \geqslant R_{e_2} \geqslant R_{e_3} \geqslant \dots ,$$

Since S satisfies $\min_{\mathcal{L}}$ and $\min_{\mathcal{R}}$, the set of $\mathcal{L}$-classes $\{L_{e_i} : i \in \mathbb{N}\}$ contains a minimal element L_{e_j} and the set of $\mathcal{R}$-classes $\{R_{e_i} : i \in \mathbb{N}\}$ contains a minimal element R_{e_k}. Let $\ell = \max\{j, k\}$; then $L_{e_\ell} = L_{e_{\ell+1}}$ and $R_{e_\ell} = R_{e_{\ell+1}}$, and so $H_{e_\ell} = H_{e_{\ell+1}}$. By Corollary 3.16, $e_\ell = e_{\ell+1}$, which is a contradiction. Thus S contains a primitive idempotent and so is completely 0-simple. 4.9

COMPLETELY SIMPLE SEMIGROUPS

Primitive idempotent

Completely simple semigroup

An idempotent of a semigroup without zero is *primitive* if it is minimal. A semigroup without zero is *completely simple* if it is simple and contains a primitive idempotent.



'Primitive idempotent' has different meanings for semigroups with and without zero: in a semigroup with a zero, a primitive idempotent is a minimal non-0 idempotent; in a semigroup without zero, a primitive idempotent is a minimal idempotent.

Finite simple $\Rightarrow$ completely simple

PROPOSITION 4.10. *A finite simple semigroup is completely simple.*

Proof of 4.10. Let S be a finite simple semigroup. Since S is finite, every element has a power which is an idempotent. So S contains at least one idempotent. Furthermore, there must be a primitive idempotent in S, since otherwise there would be an infinite descending chain of idempotents, and this is impossible since S is finite. 4.10

Define a new version of the Rees matrix construction as follows. Let G be a group, let I and Λ be abstract index sets, and let P be a $\Lambda \times I$ matrix with entries from G, with the (λ, i)-th entry of P being $p_{\lambda i}$. Let $\mathcal{M}[G; I, \Lambda; P]$ be the set $I \times G \times \Lambda$ with multiplication

$$(i, x, \lambda)(j, y, \mu) = (i, x p_{\lambda j} y, \mu).$$

Then we have the following characterization of completely simple semigroups, paralleling the Rees–Suschkewitsch theorem:

Characterization of completely simple semigroups

THEOREM 4.11. *A semigroup S is completely simple if and only if there exist a group G, index sets I and Λ, and a $\Lambda \times I$ matrix P with entries from G such that $S \simeq \mathcal{M}[G; I, \Lambda; P]$.* 4.11

Theorem 4.11, and many other properties of completely simple semigroups, are consequences of the following observations:

- S is simple if and only if S^0 is 0-simple;
- an idempotent is primitive in S if and only if it is primitive in S^0';
- for any group G, index sets I and Λ, and $\Lambda \times I$ matrix P with entries from G, we have $(\mathcal{M}[G; I, \Lambda; P])^0 = \mathcal{M}_0[G; I, \Lambda; P]$.


Notice that in the second condition above, 'primitive' means 'minimal' in S and 'minimal non-0' in S^0.

The proof of the following characterization of Green's relations in completely simple semigroups is similar to the proof of Proposition 4.8.

Green's relations in completely simple semigroups

PROPOSITION 4.12. *Let $S \simeq \mathcal{M}[G; I, \Lambda, P]$ be a completely simple semigroup.*

a) *In S, the relations $\mathcal{D}$ and $\mathcal{J}$ coincide, and S consists of a single $\mathcal{D}$-class.*

b) *The $\mathcal{L}$-classes of S are sets of the form $I \times G \times \{\lambda\}$.*

c) *The $\mathcal{R}$-classes of S are sets of the form $\{i\} \times G \times \Lambda$.*

d) *The $\mathcal{H}$-classes of S are sets of the form $\{i\} \times G \times \{\lambda\}$.* 4.12

Characterization of regular semigroups that are completely simple

PROPOSITION 4.13. *A semigroup is completely simple if and only if it is regular and every idempotent is primitive.*

Proof of 4.13. Suppose S is completely simple. Then $S \simeq \mathcal{M}[G; I, \Lambda; P]$. So S consists of a single $\mathcal{D}$-class. Furthermore, S contains idempotents, which are regular elements. Hence S is regular by Proposition 3.19. By following the reasoning in the proof of Proposition 4.2 (and ignoring mentions of the zero), every idempotent in S is primitive.

Now suppose that S is regular and every idempotent is primitive. We have to show that S is simple. Since S is regular, every $\mathcal{D}$-class contains an idempotent. So every $\mathcal{J}$-class contains an idempotent. Let J_e be a $\mathcal{J}$-class

and let $J_f \leqslant J_e$, where e and f are idempotents. Then $f = peq$ for some $p, q \in S^1$. Let $g = eqfpe$. Then

$$\begin{aligned} g^2 &= eqfpeeqfpe && \text{[by definition of } g\text{]} \\ &= eqfpeqfpe && \text{[since } e \text{ is idempotent]} \\ &= eqffpe && \text{[since } f = peq\text{]} \\ &= eqfpe && \text{[since } f \text{ is idempotent]} \\ &= g; && \text{[by definition of } g\text{]} \end{aligned}$$

thus g is idempotent. Furthermore $ge = eg = g$ and so $g \preccurlyeq e$. Since e is primitive (since all idempotents are primitive), it follows that $g = e$.

Therefore $f = peq$ and $e = g = eqfpe$. Hence $J_e = J_f$. Since all $\mathcal{J}$-classes contain idempotents, S contains only one $\mathcal{J}$-class. Hence all elements $x \in S$ generates the same principal ideal, S^1xS^1, and so $S = S^1xS^1$ for all $x \in S$. Thus S is the only ideal of S and so S is simple. 4.13

Characterization of completely simple semigroups that are groups

PROPOSITION 4.14. *A completely simple semigroup is a group if and only if it contains exactly one idempotent.*

Proof of 4.14. One direction of this result is obvious: a group contains exactly one idempotent.

Suppose S is completely simple and contains exactly one idempotent. By Proposition 4.13, S is regular. Hence, by Proposition 3.20, every $\mathcal{L}$- and every $\mathcal{R}$-class of S contains an idempotent. Since S contains only one idempotent, it contains only one $\mathcal{L}$-class and only one $\mathcal{R}$-class and so consists of a single $\mathcal{H}$-class, which is a group by Proposition 3.14. 4.14

COMPLETELY REGULAR SEMIGROUPS

Recall that a semigroup S is completely regular if it is equipped with a unary operation $^{-1}$ satisfying the conditions in (4.2); namely that for all $x \in S$,

$$(x^{-1})^{-1} = x, \qquad x^{-1}x = xx^{-1}, \qquad xx^{-1}x = x.$$

Characterization of completely regular semigroups

THEOREM 4.15. *Let S be a semigroup. Then the following are equivalent:*

a) *S is completely regular;*

b) *every element of S lies in a subgroup of S;*

c) *every $\mathcal{H}$-class of S is a subgroup.*

Proof of 4.15. Part 1 [a) $\Rightarrow$ b)]. Suppose S is completely regular. Let $x \in S$. Then $e = xx^{-1} = x^{-1}x$ is an idempotent, since $e^2 = (xx^{-1})(xx^{-1}) =$

$(xx^{-1}x)x^{-1} = xx^{-1} = e$. By Proposition 3.18, $x \in R_e \cap L_e = H_e$, which is a subgroup. So every element of S lies in a subgroup.

Part 2 [b) $\Rightarrow$ c)]. Suppose every element of S lies in a subgroup. Let $x \in S$. Then $x \in G$ for some subgroup G of S. Then $x \mathcal{H} 1_G$ and so $H_x = H_{1_G}$, which contains an idempotent and is thus a subgroup. So every $\mathcal{H}$-class of S is a subgroup.

Part 3 [c) $\Rightarrow$ a)]. Suppose every $\mathcal{H}$-class of S is a subgroup. Define $^{-1}$ by letting x^{-1} (where $x \in S$) be the unique inverse of x in the subgroup H_x. It is clear that $^{-1}$ satisfies the conditions (4.2) and S is thus completely regular. 4.15

The next result is analogous to Theorem 4.9:

THEOREM 4.16. *Let S be simple. The following are equivalent:*

Characterization of simple semigroups that are completely simple

a) *S is completely simple;*

b) *S is completely regular;*

c) *S satisfies the conditions $\min_{\mathcal{L}}$ and $\min_{\mathcal{R}}$;*

Proof of 4.16. Part 1 [a) $\Rightarrow$ b)]. Suppose S is completely simple. Then by Theorem 4.11, every element of S lies in a subgroup of S. So S is completely regular by Theorem 4.15.

Part 2 [b) $\Rightarrow$ c)]. Suppose S is completely regular. Then every element of S lies in a subgroup of S by Theorem 4.15. So every element of S^0 lies in a subgroup and so S^0 is group-bound and therefore satisfies $\min_{\mathcal{L}}$ and $\min_{\mathcal{R}}$ by Theorem 4.9.

Part 3 [c) $\Rightarrow$ a)]. Suppose S satisfies $\min_{\mathcal{L}}$ and $\min_{\mathcal{R}}$. Then so does S^0 and so S^0 is completely 0-simple by Theorem 4.9. Hence S is completely simple. 4.16

A *semilattice of semigroups* is a semigroup S for which there exists a semilattice Y and a collection of disjoint subsemigroups S_α of S, where $\alpha \in Y$, such that $S = \bigcup_{\alpha \in Y} S_\alpha$ and $S_\alpha S_\beta \subseteq S_{\alpha \sqcap \beta}$ (see Figure 4.4). A *semilattice of completely simple semigroups* is one in which every S_α is completely simple; a *semilattice of groups* is one in which every S_α is a group.

Semilattice of semigroups

THEOREM 4.17. *Every completely regular semigroup is a semilattice of completely simple semigroups.*

Characterization of completely regular semigroups

Proof of 4.17. Let S be a completely regular semigroup.

By Theorem 4.15, each $\mathcal{H}$-class of S is a subgroup. So for any $x \in S$, we have $x^2 \mathcal{H} x$ and hence $x^2 \mathcal{J} x$. Hence for any $x, y \in S$, we have $J_{xy} = J_{(xy)^2} = J_{x(yx)y} \leqslant J_{yx}$. By symmetry, $J_{yx} \leqslant J_{xy}$ and so $J_{xy} = J_{yx}$.

Let $x \mathcal{J} y$. Then there exist $r, s \in S^1$ with $rys = x$. Let $z \in S$. Then

$$J_{zx} = J_{zrys} \leqslant J_{zry} = J_{ryz} \leqslant J_{yz} = J_{zy}.$$

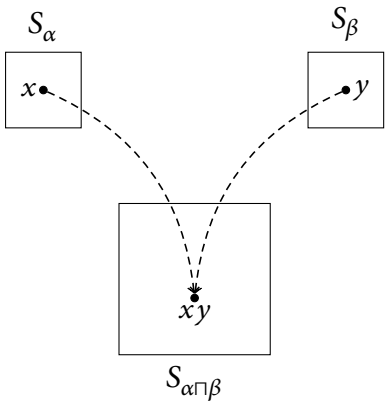


FIGURE 4.4
Multiplying in a semilattice of semigroups: the product of $x \in S_\alpha$ and $y \in S_\beta$ lies in the subsemigroup $S_{\alpha\sqcap\beta}$.

By symmetry $J_{zy} \leqslant J_{zx}$ and hence $J_{zx} = J_{zy}$. So $zx \mathcal{J} zy$. Similarly $xz \mathcal{J} yz$. Therefore $\mathcal{J}$ is a congruence. The factor semigroup $S/\mathcal{J}$ is a commutative semigroup of idempotents since $x^2 \mathcal{J} x$ and $xy \mathcal{J} yx$ for all $x, y \in S$. Hence $S/\mathcal{J}$ is a semilattice by Theorem 1.21.

Since $\mathcal{J}$ is a congruence, $J_x J_y \subseteq J_{xy}$. In particular, $J_x J_x \subseteq J_{x^2} = J_x$, so each J_x is a subsemigroup.

The aim is to show $J_x y J_x = J_x$ for all $y \in J_x$, and so deduce that J_x is simple. Let $z \in J_x$. Since $y \mathcal{J} z$, there exist $p, q, r, s \in S^1$ such that $pyq = z$ and $rzs = y$. [We cannot immediately deduce that $z \in J_x y J_x$, because p and q may not lie in J_x.] Write 1_y for the identity of H_y and 1_z for the identity of H_z. Since $y, z \in J_x$, it follows that $1_y, 1_z \in J_x$. Then $(1_z p)y(q1_z) = 1_z z 1_z = z$ and $(1_y r)z(s1_y) = 1_y y 1_y = y$. Furthermore, $J_{1_z p} \geqslant J_{(1_z p)y(q1_z)} = J_z = J_x$ and $J_{1_z p} \leqslant J_{1_z} = J_x$. Hence $1_z p \in J_x$. Similarly $q1_z \in J_x$. Hence $z = (1_z p)y(q1_z) \in J_x y J_x$. Since $z \in J_x$ was arbitrary, $J_x \subseteq J_x y J_x$. Clearly $J_x y J_x \subseteq J_x$ and so $J_x = J_x y J_x$. Since $y \in J_x$ was arbitrary, J_x is simple.

Thus, since J_x is completely regular, it is completely simple by Theorem 4.16.

To see S is a semilattice of completely simple semigroups, let Y be the semilattice $S/\mathcal{J}$ and write S_α instead of $\alpha \in S/\mathcal{J}$. 4.17

LEFT AND RIGHT GROUPS

This section discusses left and right groups, which are semigroups that are very close to being groups, and which have a very easy characterization, which we will deduce from our results on completely 0-simple semigroups.

Left/right simple semigroup

A semigroup is *left simple* if it contains no proper left ideals, and *right simple* if it contains no proper right ideals.

PROPOSITION 4.18. *Let S be left or right simple. Then S is simple.*

Proof of 4.18. Suppose S is left simple; the reasoning for the right simple case is parallel. Let $x \in S$. Then $Sx = S$ since S is left simple. So $S^2 = SS \supseteq Sx = S$ and so $S = S^2 = Sx$. Hence $SxS = S^3 = S^2 = S$ and so the only ideal of S is S itself. Thus S is simple. 4.18

Note that Proposition 4.18 shows that being left simple (or right simple) is a stronger condition than being simple. This contrasts (for example) cancellativity: being left-cancellative is a weaker condition than being cancellative.

A semigroup is a *left group* if is left simple and right cancellative, and a *right group* if it is right simple and left cancellative. Left/right group

THEOREM 4.19. *Let S be a semigroup. The following are equivalent:* Characterization of left groups

a) *S is a left group;*

b) *S is left simple and contains an idempotent;*

c) *S is completely simple semigroup and has only one $\mathcal{L}$-class;*

d) *$S \simeq Z \times G$, where Z is a left zero semigroup and G is a group.*

There is a natural analogue of Theorem 4.19 for right groups. Note that taking G trivial in part d) shows that a left zero semigroup is a left group.

Proof of 4.19. Part 1 [a) $\Rightarrow$ b)] Suppose S is a left group. By definition, S is left simple. Let $x \in S$. Since S is left simple, $Sx = S$. So there exists $e \in S$ such that $ex = x$. Thus $e^2x = ex$. Since S is right-cancellative, $e^2 = e$.

Part 2 [b) $\Rightarrow$ c)] Suppose S is left simple and $E(S) \neq \varnothing$. Since S is left simple, $S^1x = S$ for all $x \in S$, and so S consists of a single $\mathcal{L}$-class and thus a single $\mathcal{D}$-class. Since $E(S) \neq \varnothing$, some $\mathcal{H}$-class in this $\mathcal{L}$-class contains an idempotent, which is a regular element. By Proposition 3.19, all elements of S are regular. By Proposition 3.20, every $\mathcal{R}$-class of S contains an idempotent. Since S has only one $\mathcal{L}$-class, this means that every $\mathcal{H}$-class of S contains an idempotent and so is a group by Proposition 3.14. Thus S is completely regular by Theorem 4.15. Since S is left simple and thus simple by Proposition 4.18, S is completely simple by Theorem 4.16.

Part 3 [c) $\Rightarrow$ d)] Since S is completely simple, $S = \mathcal{M}[G; I, \Lambda; P]$ for some group G, index sets I and Λ, and matrix P. Since S has only one $\mathcal{L}$-class, $\Lambda = \{1\}$ by Proposition 4.12.

Make I a left zero semigroup by defining $ij = i$ for all $i, j \in I$. Define a map

$$\varphi : I \times G \to S; \qquad (i, g)\varphi = (i, p_{1i}^{-1} g, 1).$$

Note that in this definition, the pair (i, g) is in the direct product $I \times G$, and the triple $(i, p_{1i}^{-1} g, 1)$ is in the Rees matrix semigroup $\mathcal{M}[G; I, \Lambda; P] = S$.

Then

$$\begin{aligned}
&((i,g)\varphi)((j,h)\varphi)\\
&= (i, p_{1i}^{-1}g, 1)(j, p_{1j}^{-1}h, 1) && \text{[by definition of } \varphi]\\
&= (i, p_{1i}^{-1}gp_{1j}p_{1j}^{-1}h, 1) && \text{[by multiplication in } \mathcal{M}[G;I,\Lambda;P]]\\
&= (i, p_{i1}^{-1}gh, 1) && \text{[by multiplication in } G]\\
&= (i,gh)\varphi && \text{[by definition of } \varphi]\\
&= \big((i,g)(j,h)\big)\varphi, && \text{[by multiplication in } I\times G]
\end{aligned}$$

so φ is a homomorphism.

Furthermore,

$$\begin{aligned}
(i,g)\varphi = (j,h)\varphi &\Rightarrow (i, p_{1i}^{-1}g, 1) = (j, p_{1j}^{-1}h, 1) && \text{[by definition of } \varphi]\\
&\Rightarrow i = j \wedge p_{1i}^{-1}g = p_{1j}^{-1}h\\
&\Rightarrow i = j \wedge p_{1i}^{-1}g = p_{1i}^{-1}h && \text{[using } i=j]\\
&\Rightarrow i = j \wedge g = h && \text{[by cancellation in } G]\\
&\Rightarrow (i,g) = (j,h);
\end{aligned}$$

thus φ is injective.

Finally, for any $(i,g,1) \in S = \mathcal{M}[G;I,\Lambda;P]$, we have $(i,p_{1i}g)\varphi = (i, p_{1i}^{-1}p_{1i}g, 1) = (i,g,1)$, so φ is surjective. So φ is an isomorphism, and $S \simeq I\times G$. Since I is a left zero semigroup and G is a group, this completes this part of the proof.

Part 4 [d) $\Rightarrow$ a)] Let $(x,g),(y,h),(z,i) \in Z\times G$. Then

$$\begin{aligned}
&(x,g)(y,h) = (x,g)(z,i)\\
&\Rightarrow (y,gh) = (z,gi) && \text{[since } Z \text{ is a right zero semigroup]}\\
&\Rightarrow (y=z)\wedge(gh=gi)\\
&\Rightarrow (y=z)\wedge(h=i) && \text{[since } G \text{ is a group]}\\
&\Rightarrow (y,h) = (z,i).
\end{aligned}$$

So $Z\times G$ is right-cancellative. Furthermore, $(x,g)(y,g^{-1}h) = (y,h)$ and so $(g,x)(Z\times G) = Z\times G$ for all $(x,g) \in Z\times G$. Hence $Z\times G$ is left simple, and so $Z\times G$ is a left group. 4.19

HOMOMORPHISMS

We close this chapter with the following result, showing that homomorphisms preserve regularity and that, within regular semigroups, the preimage of an idempotent must contain an idempotent.

PROPOSITION 4.20. *Let S be a regular semigroup, T a semigroup (not necessarily regular), and let $\varphi : S \to T$ be a homomorphism.*

a) *The subsemigroup* $\operatorname{im}\varphi$ *of T is a regular semigroup and that if $x' \in S$ is an inverse of $x \in S$, then $x'\varphi$ is an inverse of $x\varphi$.*

b) *If $e \in \operatorname{im}\varphi$ is idempotent, $f\varphi = e$, and $z \in S$ is an inverse of f^2, then fzf is idempotent and $(fzf)\varphi = e$.*

Proof of 4.20. a) Clearly $\operatorname{im}\varphi$ is a semigroup; we have to show it is inverse. Let $y \in \operatorname{im}\varphi$. Then there exists $x \in S$ with $x\varphi = y$. Since S is regular, there exists an inverse x' for x. Let $y' = x'\varphi$. Then $yy'y = (x\varphi)(x'\varphi)(x\varphi) = (xx'x)\varphi = x\varphi = y$ and similarly $y'yy' = y$. So y' is an inverse for y. Since $y \in \operatorname{im}\varphi$ was arbirary, $\operatorname{im}\varphi$ is regular.

b) Let $g = fzf$. Since z is an inverse of f^2, we have $zf^2z = z$ and $f^2zf^2 = f^2$. Then $g^2 = f(zf^2z)f = fzf = g$ and so g is idempotent. Furthermore

$$\begin{aligned}
g\varphi &= (fzf)\varphi && \text{[by choice of } g\text{]}\\
&= (f\varphi)(z\varphi)(f\varphi) && \text{[since } \varphi \text{ is a homomorphism]}\\
&= e(z\varphi)e && \text{[since } e = f\varphi\text{]}\\
&= e^2(z\varphi)e^2 && \text{[since } e \text{ is idempotent]}\\
&= (f\varphi)^2(z\varphi)(f\varphi)^2 && \text{[since } e = f\varphi\text{]}\\
&= (f^2zf^2)\varphi && \text{[since } \varphi \text{ is a homomorphism]}\\
&= (f^2\varphi) && \text{[since } f^2zf^2 = f^2\text{]}\\
&= (f\varphi)^2 && \text{[since } \varphi \text{ is a homomorphism]}\\
&= e^2 && \text{[since } f\varphi = e\text{]}\\
&= e. && \text{[since } e \text{ is idempotent]}
\end{aligned}$$

This completes the proof. 4.20

EXERCISES

[See pages 220–225 for the solutions.]

4.1 Let G be a group, let $I = \{1\}$ and $\Lambda = \{1\}$ be index sets (each containing only one element), and P a $\Lambda \times I$ matrix over G.

a) By defining a suitable isomorphism, prove that $\mathcal{M}[G; I, \Lambda; P] \simeq G$.

b) Give an example to show that if we replace G by a monoid M and construct $\mathcal{M}[M; I, \Lambda; P]$ using the same multiplication, we can have $\mathcal{M}[M; I, \Lambda; P] \neq M$.

4.2 Prove that every completely simple semigroup is equidivisible.

4.3 Let S be a completely simple semigroup. Prove that

a) $\mathcal{L}$, $\mathcal{R}$, and $\mathcal{H}$ are congruences on S;

b) $S/\mathcal{L}$ is a right zero semigroup and $S/\mathcal{R}$ is a left zero semigroup;

c) $S/\mathcal{H}$ is isomorphic to the rectangular band $S/\mathcal{R} \times S/\mathcal{L}$.

4.4 Let S be a completely simple semigroup.

a) Suppose $|S| = p$, where p is a prime. Prove that S is [isomorphic to] either a right zero semigroup, a left zero semigroup, or a group.

b) Suppose $|S| = pq$, where p and q are primes. Prove that S is [isomorphic to] either a rectangular band, a right group, or a left group.

∗4.5 a) Let S and T be completely regular semigroups and $\varphi : S \to T$ a homomorphism. Show that $(z\varphi)^{-1} = z^{-1}\varphi$ for all $z \in S$.

b) Give an example of regular semigroups S and T that have operations $^{-1}$ satisfying (4.1), and a homomorphism $\varphi : S \to T$ such that $(z\varphi)^{-1} \neq z^{-1}\varphi$ for some $z \in S$.

∗4.6 Let G and H be groups, I, J, Λ, and M be index sets, P be a $\Lambda \times I$ regular matrix over G^0, and Q be a $J \times M$ regular matrix over H^0.

a) Suppose $\varphi : \mathcal{M}_0[G; I, \Lambda; P] \to \mathcal{M}_0[H; J, M; Q]$ is an isomorphism.

i) Prove that there exist bijections $\alpha : I \to J$ and $\beta : \Lambda \to M$ such that $(i, a, \lambda)\varphi \in \{i\alpha\} \times H \times \{\lambda\beta\}$ and such that $p_{\lambda i} = 0 \Leftrightarrow q_{(\lambda\beta)(i\alpha)} = 0$.

ii) Assume without loss that $1 \in I \cap \Lambda$. Define an isomorphism $\gamma : G \to \{1\} \times G \times \{1\} \subseteq \mathcal{M}_0[G; I, \Lambda; P]$ and an isomorphism $\eta : H \to \{1\alpha\} \times H \times \{1\beta\} \subseteq \mathcal{M}_0[H; J, M; Q]$. Deduce that $\vartheta = \gamma\varphi\eta^{-1}$ is an isomorphism from G to H.

iii) Check that $(i, x, \lambda) = (i, 1_G, 1)(1, p_{11}^{-1}x, 1)(1, p_{11}^{-1}, \lambda)$. Now let $u_i, v_\lambda \in H$ be such that

$$(i, 1_G, 1)\varphi = (i\alpha, u_i, 1\beta)$$

and

$$(1, p_{11}^{-1}, \lambda)\varphi = (1\alpha, q_{(1\alpha)(1\beta)}^{-1}v_\lambda, \lambda\beta).$$

Using the fact that $(i, p_{\lambda i}, \lambda) = (i, 1_G, \lambda)(i, 1_G, \lambda)$, prove that

$$p_{\lambda i}\vartheta = v_\lambda q_{(\lambda\beta)(i\alpha)}u_i \tag{4.5}$$

for all $i \in I$ and $\lambda \in \Lambda$.

b) Suppose that there exists an isomorphism $\vartheta : G \to H$, bijections $\alpha : I \to J$ and $\beta : \Lambda \to M$ and elements u_i and v_λ such that (4.5) holds for all $i \in I$ and $\lambda \in \Lambda$. Show that $\mathcal{M}_0[G; I, \Lambda; P] \simeq \mathcal{M}_0[H; J, M; Q]$.

✶4.7 Let G be a group, I and Λ index sets, and let P be a $\Lambda \times I$ matrix over G^0 that is not necessarily regular (that is, P can have rows or columns where all the entries are 0). Let $S = \mathcal{M}_0[G; I, \Lambda; P]$, where the multiplication is the same as in the usual Rees matrix semigroup. Prove that S is regular (as a semigroup) if and only if P is regular (as a matrix).

4.8 Let S be a 0-simple semigroup.

a) Suppose S satisfies $\min_{\mathcal{L}}$. By applying this condition to the set of $\mathcal{L}$-classes that are not equal to $\{0\}$, show that S contains a 0-minimal left ideal. [Dually, if S satisfies $\min_{\mathcal{R}}$, it contains a 0-minimal right ideal.]

b) Now suppose S that S contains a 0-minimal left ideal and a 0-minimal right ideal.

i) Prove that if K is a 0-minimal left ideal of S with $K^2 \neq \{0\}$, then $K = Sx$ for any $x \in K \setminus \{0\}$.

ii) Let L be a 0-minimal left ideal of S, and suppose $x \in S$ is such that $Lx \neq \{0\}$. Prove that Lx is a 0-minimal left ideal of S. [Hint: to prove Lx is 0-minimal, consider the set $J = \{ y \in L : yx \in K \}$.]

iii) By considering the subset LS of S, prove that there exists $x \in S$ with $Lx \neq \{0\}$.

iv) Let $M = \bigcup\{ Lx : x \in S, Lx \neq \{0\} \}$. Prove that $M = S$ and deduce that S is the union of its 0-minimal left ideals. [Dually, S is the union of its 0-minimal right ideals.]

v) Using part iii), the dual version of part iv), and Exercise 1.21, prove that there exists a 0-minimal right ideal R such that $LR = S$ and RL is a group with a zero adjoined.

vi) Let e be the identity of the group $RL \setminus \{0\}$. Prove that e is a primitive idempotent.

This proves that S is completely 0-simple.

Since a completely 0-simple semigroup satisfies $\min_{\mathcal{L}}$ and $\min_{\mathcal{R}}$ by Theorem 4.9, this exercise has shown that a 0-simple semigroup is completely 0-simple if and only it has a 0-minimal left ideal and a 0-minimal right ideal.

4.9 Let S be a left-cancellative semigroup. Let G be a subgroup of S. Suppose G is also a left ideal of S. Prove that S is a right group.

NOTES

The exposition here is based on Howie, *Fundamentals of Semigroup Theory*, ch. 3 and Clifford & Preston, *The Algebraic Theory of Semigroups*,

§ 2.5. ◆ The Rees–Suschkewitsch theorem (Theorem 4.7) was originally proved in Rees & Hall, 'On semi-groups'; the analogue for completely simple semigroups (Theorem 4.11) is the earlier version, having been essentially proved in Suschkewitsch, 'Über die endlichen Gruppen…' ◆ The results on the structure of completely regular semigroups are due to Clifford, 'Semigroups admitting relative inverses'. ◆ The analogy of Theorem 4.19 for right groups is due to Suschkewitsch, 'Über die endlichen Gruppen…'; for a more accessible proof (which does not use Green's relations or the Rees–Suschkewitsch theorem), see Clifford & Preston, *The Algebraic Theory of Semigroups*, Theorem 1.27 ◆ Exercise 4.9 is from Cain, Robertson & Ruškuc, 'Cancellative and Malcev presentations for finite Rees index subsemigroups and extensions', Proposition 8.3. ◆ For further reading, Petrich, *Completely Regular Semigroups* seems to be the most recent monograph in the area.

Inverse semigroups

5

> ‘The sensibility of man to trifles, and his insensibility to great things, indicates a strange inversion.’
>
> — Blaise Pascal, *Pensées*, § iii.198.

Inverse semigroup

❁ Recall that an inverse semigroup is one equipped with an operation $^{-1}$ satisfying the four conditions in (4.3): namely, that for all $x, y \in S$,

$$(x^{-1})^{-1} = x, \tag{5.1}$$
$$(xy)^{-1} = y^{-1}x^{-1}, \tag{5.2}$$
$$xx^{-1}x = x, \tag{5.3}$$
$$xx^{-1}yy^{-1} = yy^{-1}xx^{-1}. \tag{5.4}$$

Clifford & Preston’s 1961 view that ‘[i]nverse semigroups constitute probably the most promising class of semigroups for study’ has proved accurate, and the field has grown into a vast and active one. We can only survey a minuscule part of it here.

EQUIVALENT CHARACTERIZATIONS

We being by giving alternative characterizations of inverse semigroups. Some texts define inverse semigroups using one of these alternative characterizations.

Characterizations of inverse semigroups

THEOREM 5.1. *The following are equivalent:*

a) *S is an inverse semigroup;*

b) *every element of S has a unique inverse;*

c) *S is regular and its idempotents commute;*

d) *every $\mathcal{L}$-class and every $\mathcal{R}$-class of S contains exactly one idempotent.*

Proof of 5.1. The plan is as follows: parts 1–3 of this proof show that b), c), and d) are equivalent. Then parts 4 and 5 show, respectively, that a) implies c) and that b) implies a).

Part 1 [b) ⇒ c)]. Suppose every element of S has a unique inverse. Then S is clearly regular. Let $e, f \in E(S)$. Then

$$\begin{aligned}
&(ef)(f(ef)^{-1}e)(ef) \\
&= ef^2(ef)^{-1}e^2f && \text{[rearranging brackets]} \\
&= ef(ef)^{-1}ef && \text{[since } e \text{ and } f \text{ are idempotent]} \\
&= ef && \text{[by definition of inverse]}
\end{aligned}$$

and

$$\begin{aligned}
&(f(ef)^{-1}e)(ef)(f(ef)^{-1}e) \\
&= f(ef)^{-1}e^2f^2(ef)^{-1}e && \text{[rearranging brackets]} \\
&= f(ef)^{-1}ef(ef)^{-1}e && \text{[since } e \text{ and } f \text{ are idempotent]} \\
&= f(ef)^{-1}e && \text{[by definition of inverse]}
\end{aligned}$$

and so $f(ef)^{-1}e$ is an inverse of ef. Since inverses are unique, $(ef)^{-1} = f(ef)^{-1}e$. Hence

$$\begin{aligned}
&((ef)^{-1})^2 \\
&= f(ef)^{-1}ef(ef)^{-1}e && \text{[since } (ef)^{-1} = f(ef)^{-1}e\text{]} \\
&= f(ef)^{-1}e && \text{[by definition of inverse]} \\
&= (ef)^{-1} && \text{[since } (ef)^{-1} = f(ef)^{-1}e\text{]}
\end{aligned}$$

and so $(ef)^{-1}$ is idempotent. Thus $(ef)^{-1}(ef)^{-1}(ef)^{-1} = (ef)^{-1}$ and so the uniqueness of inverses implies that $ef = ((ef)^{-1})^{-1} = (ef)^{-1}$ and so ef is idempotent. A similar argument shows that fe is idempotent. Hence

$$(ef)(fe)(ef) = ef^2e^2f = efef = ef$$

and

$$(fe)(ef)(fe) = fe^2f^2e = fefe = fe.$$

Hence $fe = (ef)^{-1} = ef$. Thus idempotents of S commute.

Part 2 [c) ⇒ d)]. Suppose that S is regular and that its idempotents commute. Since S is regular, every $\mathcal{L}$-class contains at least one idempotent by Proposition 3.20. So suppose a particular $\mathcal{L}$-class contains idempotents e and f. Then both e and f are right identities for this $\mathcal{L}$-class by Proposition 3.17. So $ef = e$ and $fe = f$. Since idempotents commute, $ef = fe$ and so $e = f$. So each $\mathcal{L}$-class contains a unique idempotent. Similarly each $\mathcal{R}$-class contains a unique idempotent.

Part 3 [d) ⇒ b)]. Suppose every $\mathcal{L}$-class and every $\mathcal{R}$-class of S contains a unique idempotent. Let $x \in S$. By Proposition 3.21, the inverses of x are in one-to-one correspondence with pairs of idempotents $(e, f) \in R_x \times L_x$.

Since R_x and L_x each contain a unique idempotent, x therefore has a unique inverse. So every element of S has a unique inverse.

Part 4 [a) $\Rightarrow$ c)]. Suppose S is an inverse semigroup. Let $x \in S$. Then $xx^{-1}x = x$ by (5.3) and so S is regular. Let $e \in E(S)$. Then

$$\begin{aligned}
e^{-1} &= e^{-1}(e^{-1})^{-1}e^{-1} && \text{[by (5.3)]}\\
&= e^{-1}ee^{-1} && \text{[by (5.1)]}\\
&= e^{-1}eee^{-1} && \text{[since } e \text{ is idempotent]}\\
&= e^{-1}(e^{-1})^{-1}ee^{-1} && \text{[by (5.1)]}\\
&= ee^{-1}e^{-1}(e^{-1})^{-1} && \text{[by (5.4)]}\\
&= ee^{-1}e^{-1}e && \text{[by (5.1)]}\\
&= e(ee)^{-1}e && \text{[by (5.2)]}\\
&= ee^{-1}e && \text{[since } e \text{ is idempotent]}\\
&= e. && \text{[by (5.3)]}
\end{aligned}$$

Hence $ee^{-1} = e^2 = e$ and $e^{-1}e = e^2 = e$ for any $e \in E(S)$. Now let $e, f \in E(S)$. Then $ef = ee^{-1}ff^{-1} = ff^{-1}ee^{-1} = fe$ by (5.4). Thus idempotents of S commute.

Part 5 [b) $\Rightarrow$ a)]. Suppose every element of S has a unique inverse. Then for any $x \in S$, we have $xx^{-1}x = x$; thus (5.3) holds. By the uniqueness of inverses, $(x^{-1})^{-1} = x$; thus (5.1) holds. Let $x, y \in S$. Then xx^{-1} and yy^{-1} are idempotents and so commute by parts 1 and 2 of this proof; thus (5.4) holds. Therefore

$$\begin{aligned}
&xy(y^{-1}x^{-1})xy\\
&= x(yy^{-1})(x^{-1}x)y && \text{[rearranging brackets]}\\
&= xx^{-1}xyy^{-1}y && \text{[by (5.4), which holds]}\\
&= xy && \text{[by definition of inverse]}
\end{aligned}$$

and

$$\begin{aligned}
&(y^{-1}x^{-1})xy(y^{-1}x^{-1})\\
&= y^{-1}(x^{-1}x)(yy^{-1})x^{-1} && \text{[rearranging brackets]}\\
&= y^{-1}yy^{-1}x^{-1}xx^{-1} && \text{[by (5.4), which holds]}\\
&= y^{-1}x^{-1}. && \text{[by definition of inverse]}
\end{aligned}$$

Hence, by the uniqueness of inverses, $(xy)^{-1} = y^{-1}x^{-1}$; thus (5.2) holds. Thus S is an inverse semigroup. 5.1

We now prove some consequences of these alternative characterizations.

PROPOSITION 5.2. *Let S be an inverse semigroup. Then $E(S)$ is a subsemigroup of S and forms a semilattice.*

$E(S)$ is a semilattice when S is inverse

Proof of 5.2. By Theorem 5.1, all elements of $E(S)$ commute. Hence, if $e, f \in E(S)$, then $(ef)^2 = efef = e^2 f^2 = ef$ and so $ef \in E(S)$. So $E(S)$ is a subsemigroup of S. Furthermore, $E(S)$ is a commutative semigroup of idempotents and hence a semilattice by Theorem 1.21. 5.2

Characterization of inverse semigroups that are groups

PROPOSITION 5.3. *Let S be an inverse semigroup. Then S is a group if and only if S contains exactly one idempotent.*

Proof of 5.3. In one direction, this result is obvious: if S is a group, then it is an inverse semigroup and 1_S is the unique idempotent in S.

So suppose S is an inverse semigroup and e is the unique idempotent in S. Let $x \in S$. Then xx^{-1} and $x^{-1}x$ are idempotents and so $e = xx^{-1} = x^{-1}x$. Thus $ex = xx^{-1}x = x$ and $xe = xx^{-1}x = x$. So e is an identity for S. Furthemore, since $e = xx^{-1} = x^{-1}x$ for all $x \in S$, every element of x is right- and left-invertible and so S is a group. 5.3

Inverse semigroup homomorphism

Let S and T be inverse semigroups. A homomorphism $\varphi : S \to T$ is an *inverse semigroup homomorphism* if $x^{-1}\varphi = (x\varphi)^{-1}$ for all $x \in S$.

In Chapter 1, we saw the distinction between a [semigroup] homomorphism and a monoid homomorphism: a homomorphism between two monoids may preserve multiplication, but not preserve the identity (see Exercise 1.15). Thus it is conceivable that there exists a homomorphism between two inverse semigroups that is not an inverse semigroup homomorphism. However, the following result shows that the two notions coincide:

Homomorphism from an inverse semigroup is an inverse semigroup homomorphism

PROPOSITION 5.4. *Let S be an inverse semigroup and let T a semigroup (not necessarily inverse), and let $\varphi : S \to T$ be a homomorphism. Then $\operatorname{im}\varphi$ is an inverse semigroup, and φ is an inverse semigroup homomorphism.*

Proof of 5.4. By Proposition 4.20(a), $\operatorname{im}\varphi$ is regular and $x^{-1}\varphi$ is an inverse of $x\varphi$ for any $x \in S$. Let $e, f \in \operatorname{im}\varphi$ be idempotents. By Proposition 4.20(b), there are idempotents $g, h \in S$ with $g\varphi = e$ and $h\varphi = f$. Since S is an inverse semigroup, $gh = hg$ by Theorem 5.1. Thus $ef = (g\varphi)(h\varphi) = (gh)\varphi = (hg)\varphi = (h\varphi)(g\varphi) = fe$. Hence idempotents commute in $\operatorname{im}\varphi$ and so $\operatorname{im}\varphi$ is an inverse semigroup by Theorem 5.1. Since inverses are unique in inverse semigroups, it follows that $(x\varphi)^{-1} = x^{-1}\varphi$ for all $x \in S$, so φ is an inverse semigroup homomorphism. 5.4

Homomorphism from a group preserves identity and inverses

COROLLARY 5.5. *Let G be a group and let T a semigroup (not necessarily a group or inverse), and let $\varphi : G \to T$ be a homomorphism. Then $\operatorname{im}\varphi$ is a group, and $\varphi : G \to \operatorname{im}\varphi$ is an inverse semigroup homomorphism and a monoid homomorphism. (That is, $x^{-1}\varphi = (x\varphi)^{-1}$ for all $x \in G$ and $1_G\varphi$ is an identity for $\operatorname{im}\varphi$.)*

Proof of 5.5. Proposition 5.4 shows that $\operatorname{im}\varphi$ is an inverse semigroup and φ is an inverse semigroup homomorphism. Let $y \in \operatorname{im}\varphi$ and let $x \in G$ be

such that $x\varphi = y$. Then $(1_G\varphi)y = ((xx^{-1})\varphi)(x\varphi) = (xx^{-1}x)\varphi = x\varphi = y$, and similarly $y(1_G\varphi) = y$. Hence $1_G\varphi$ is an identity for $\operatorname{im}\varphi$. 5.5

The last consequence we prove is more technical, but we will make use of it in the next section.

LEMMA 5.6. *Let S be an inverse semigroup.*

a) *For any $e, f \in E(S)$, we have $Se = Sf \Rightarrow e = f$.*

b) *For any $e, f \in E(S)$, we have $Se \cap Sf = Sef$.*

c) *For any $x \in S$, we have $Sx = Sx^{-1}x$.*

d) *For $x \in S$ and $e \in E(S)$, the element $f = x^{-1}ex$ is idempotent and $ex = xf$.*

Proof of 5.6. a) Since $e = ee \in Se = Sf$, we deduce that $e = xf$ for some $x \in S$. Then $ef = xf^2 = xf = e$. Similarly $fe = f$. Since idempotents commute by Theorem 5.1, $e = f$.

b) Obviously $Sef \subseteq Sf$ and, since idempotents commute, $Sef = Sfe \subseteq Se$. So $Sef \subseteq Se \cap Sf$. Let $x \in Se \cap Sf$. Then $x = ye$ and $x = zf$ for some $y, z \in S$. Then $x = zf = zf^2 = xf = yef \in Sef$. So $Se \cap Sf \subseteq Sef$ and hence $Se \cap Sf = Sef$.

c) Obviously $Sx^{-1}x \subseteq Sx$. But $Sx = Sxx^{-1}x \subseteq Sx^{-1}x$ and so $Sx = Sx^{-1}x$.

d) Since xx^{-1} is an idempotent, and idempotents commute in S, $f^2 = x^{-1}exx^{-1}ex = x^{-1}xx^{-1}eex = x^{-1}ex = f$, so f is idempotent. Furthermore, $ex = exx^{-1}x = xx^{-1}ex = xf$. 5.6

VAGNER–PRESTON THEOREM

Theorem 1.22 showed that every semigroup embeds into $\mathcal{T}_X$ for some X. Cayley's theorem shows that every group embeds into $\mathcal{S}_X$ for some X. The Vagner–Preston theorem, to which this section is devoted, is an analogue of these results for inverse semigroups.

Let $\tau \in \mathcal{P}_X$. Recall from (1.3) that the domain of τ, denoted $\operatorname{dom}\tau$, is the subset of X on which τ is defined. If $\tau : \operatorname{dom}\tau \to \operatorname{im}\tau$ is a bijection, then τ is a *partial bijection*. The set of partial bijections on X is denoted $\mathcal{I}_X$. (The symbol $\mathcal{I}$ stands for 'injection'.) Notice that if $\tau, \sigma \in \mathcal{I}_X$, then

Partial bijection

$\mathcal{I}_X$

$$\begin{aligned} x \in \operatorname{dom}(\tau\sigma) &\Leftrightarrow (\exists y \in X)((x,y) \in \tau\sigma) \\ &\Leftrightarrow (\exists y \in X)(\exists z \in X)\big((x,z) \in \tau \wedge (z,y) \in \sigma\big) \\ &\Leftrightarrow (\exists z \in X)\big((x,z) \in \tau \wedge z \in \operatorname{dom}\sigma\big) \\ &\Leftrightarrow (x \in \operatorname{dom}\tau) \wedge (x\tau \in \operatorname{dom}\sigma) \\ &\Leftrightarrow (x\tau \in \operatorname{im}\tau) \wedge (x\tau \in \operatorname{dom}\sigma) \end{aligned}$$

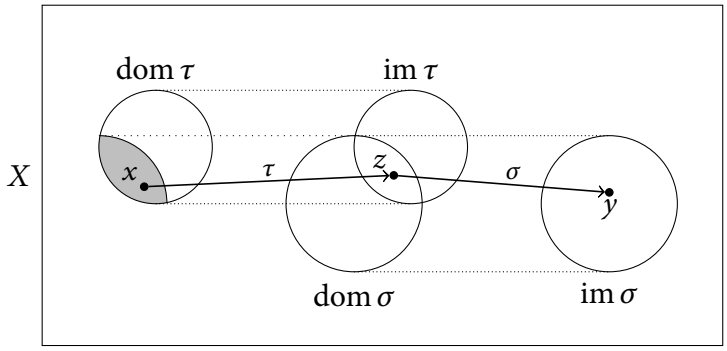


FIGURE 5.1
The domain of the composition of two partial bijections τ and σ; the shaded area is $\mathrm{dom}(\tau\sigma)$.

$$\Leftrightarrow (x\tau \in \mathrm{im}\,\tau \cap \mathrm{dom}\,\sigma)$$
$$\Leftrightarrow x \in (\mathrm{im}\,\tau \cap \mathrm{dom}\,\sigma)\tau^{-1}.$$

That is,

$$\mathrm{dom}(\tau\sigma) = (\mathrm{im}\,\tau \cap \mathrm{dom}\,\sigma)\tau^{-1}. \tag{5.5}$$

(See Figure 5.1.) For $x, y \in \mathrm{dom}\,\tau\sigma$, we have $x, y \in \mathrm{dom}\,\tau$ and $x\tau, y\tau \in \mathrm{dom}\,\sigma$ and so

$$\begin{aligned} x\tau\sigma = y\tau\sigma &\Rightarrow x\tau = y\tau && \text{[since } \sigma \text{ is injective]} \\ &\Rightarrow x = y. && \text{[since } \tau \text{ is injective]} \end{aligned}$$

Hence $\tau\sigma$ is a bijection from $\mathrm{dom}(\tau\sigma)$ to $\mathrm{im}(\tau\sigma)$ and so $\tau\sigma \in \mathcal{I}_X$. Thus $\mathcal{I}_X$ is a subsemigroup of $\mathcal{P}_X$.

Inverse of a partial bijection

For $\tau \in \mathcal{I}_X$, let τ^{-1} be the partial bijection with domain $\mathrm{im}\,\tau$ and image $\mathrm{dom}\,\tau$ defined by $(x\tau)\tau^{-1} = x$. (That is, τ is defined by inverting the bijection $\tau : \mathrm{dom}\,\tau \to \mathrm{im}\,\tau$.) Note that

$$\tau\tau^{-1} = \mathrm{id}_{\mathrm{dom}\,\tau} \quad \text{and} \quad \tau^{-1}\tau = \mathrm{id}_{\mathrm{im}\,\tau}. \tag{5.6}$$

$\mathcal{I}_X$ is an inverse semigroup

PROPOSITION 5.7. *For any set X, the semigroup of partial bijections $\mathcal{I}_X$ is an inverse semigroup.*

Proof of 5.7. Let $\tau \in \mathcal{I}_X$. Since $\tau\tau^{-1} = \mathrm{id}_{\mathrm{dom}\,\tau}$ and $\tau^{-1}\tau = \mathrm{id}_{\mathrm{im}\,\tau}$ by (5.6), we have $\tau\tau^{-1}\tau = \tau$ and $\tau^{-1}\tau\tau^{-1} = \tau^{-1}$. Hence τ^{-1} is an inverse of τ. Thus $\mathcal{I}_X$ is regular.

Suppose $\sigma \in \mathcal{I}_X$ is an inverse of τ. Then $\tau\sigma\tau = \tau$ and $\sigma\tau\sigma = \sigma$. Suppose, with the aim of obtaining a contradiction, that $\mathrm{im}\,\tau \not\subseteq \mathrm{dom}(\sigma\tau)$. So there exists $t \in \mathrm{im}\,\tau \setminus \mathrm{dom}(\sigma\tau)$; thus $t \in \mathrm{im}\,\tau$ but $t \notin \mathrm{im}\,\tau \cap \mathrm{dom}(\sigma\tau)$. Hence $\mathrm{im}\,\tau \cap \mathrm{dom}(\sigma\tau) \subsetneq \mathrm{im}\,\tau$. Therefore

$$\begin{aligned} \mathrm{dom}\,\tau &= \mathrm{dom}(\tau\sigma\tau) \\ &= (\mathrm{im}\,\tau \cap \mathrm{dom}(\sigma\tau))\tau^{-1} && \text{[by (5.5)]} \\ &\subsetneq (\mathrm{im}\,\tau)\tau^{-1} \\ &= \mathrm{dom}\,\tau. \end{aligned}$$

The strict inclusion is a contradiction; hence $\operatorname{im}\tau \subseteq \operatorname{dom}(\sigma\tau) \subseteq \operatorname{dom}\sigma$. Similarly, from $\sigma\tau\sigma = \sigma$ we obtain $\operatorname{dom}\sigma \subseteq \operatorname{im}\tau$. Therefore $\operatorname{dom}\sigma = \operatorname{im}\tau = \operatorname{dom}\tau^{-1}$. For any $x \in \operatorname{dom}\sigma$, we have $x \in \operatorname{im}\tau$ and so $x = y\tau$ for some $y \in X$. Hence $x\sigma = y\tau\sigma = y\tau\sigma\tau\tau^{-1} = y\tau\tau^{-1} = x\tau^{-1}$. Hence $\sigma = \tau^{-1}$. So τ^{-1} is the unique inverse of τ.

Since each element of $\mathcal{I}_X$ has a unique inverse, $\mathcal{I}_X$ is an inverse semigroup by Theorem 5.1. 5.7

Let S be an inverse semigroup and let T be a subsemigroup of S. Then T is an *inverse subsemigroup* of S if it is also an inverse semigroup, or, equivalently, if it is closed under taking inverses in S. Inverse subsemigroup

VAGNER–PRESTON THEOREM 5.8. *For any inverse semigroup S, there exists a set X and a monomorphism $\varphi : S \to \mathcal{I}_X$. Hence every inverse semigroup is isomorphic to some inverse subsemigroup of $\mathcal{I}_X$.* Vagner–Preston theorem

Proof of 5.8. Let $X = S$. For each $x \in S$, let τ_x be the partial transformation with domain Sx^{-1} and defined by $y\tau_x = yx$. Thus τ_x is simply ρ_x (as defined on page 19) restricted to Sx^{-1}. Note that $\operatorname{im}\tau_x = (\operatorname{dom}\tau_x)\tau_x = Sx^{-1}x = Sx$, by Lemma 5.6(c).

Let us prove that $\tau_x \in \mathcal{I}_X$. Let $y, z \in Sx^{-1}$, with $y = px^{-1}$ and $z = qx^{-1}$. Then

$$\begin{aligned} y\tau_x = z\tau_x &\Rightarrow yx = zx \\ &\Rightarrow px^{-1}x = qx^{-1}x \\ &\Rightarrow px^{-1}xx^{-1} = qx^{-1}xx^{-1} \\ &\Rightarrow px^{-1} = qx^{-1} \\ &\Rightarrow y = z. \end{aligned}$$

So τ_x is a partial bijection and so $\tau_x \in \mathcal{I}_X$.

Let us now prove that $(\tau_x)^{-1} = \tau_{x^{-1}}$. If $z \in \operatorname{dom}\tau_x = Sx^{-1}$, then $z\tau_x\tau_{x^{-1}}\tau_x = zxx^{-1}x = zx = z\tau_x$. If $z \in \operatorname{dom}\tau_{x^{-1}} = Sx$, then $z\tau_{x^{-1}}\tau_x\tau_{x^{-1}} = zx^{-1}xx^{-1} = zx^{-1} = z\tau_{x^{-1}}$. Furthermore, $\operatorname{dom}\tau_{x^{-1}} = Sx = \operatorname{im}\tau_x$ and $\operatorname{im}\tau_{x^{-1}} = Sx^{-1} = \operatorname{dom}\tau_x$. Hence $(\tau_x)^{-1} = \tau_{x^{-1}}$.

Define $\varphi : S \to \mathcal{I}_X$ by $x\varphi = \tau_x$. We first prove that φ is injective. Let $x, y \in S$. Then

$$\begin{aligned} x\varphi = y\varphi &\Rightarrow \tau_x = \tau_y && \text{[by definition of } \varphi\text{]} \\ &\Rightarrow \operatorname{dom}\tau_x = \operatorname{dom}\tau_y \\ &\Rightarrow Sx^{-1} = Sy^{-1} && \text{[by definition of } \tau_x \text{ and } \tau_y\text{]} \\ &\Rightarrow Sxx^{-1} = Syy^{-1} && \text{[by Lemma 5.6(c)]} \\ &\Rightarrow xx^{-1} = yy^{-1} && \text{[by Lemma 5.6(a)]} \\ &\Rightarrow xx^{-1}\tau_x = yy^{-1}\tau_y && \text{[since } \tau_x = \tau_y\text{]} \\ &\Rightarrow xx^{-1}x = yy^{-1}y && \text{[by definition of } \tau_x \text{ and } \tau_y\text{]} \\ &\Rightarrow x = y; \end{aligned}$$

thus φ is injective.

Let $x, y \in S$. Then

$$\begin{aligned}
\operatorname{dom}(\tau_x\tau_y) &= (\operatorname{im}\tau_x \cap \operatorname{dom}\tau_y)\tau_x^{-1} && \text{[by (5.5)]}\\
&= (Sx^{-1}x \cap Sy^{-1})\tau_x^{-1} && \text{[by definition of } \tau_x \text{ and } \tau_y\text{]}\\
&= (Sx^{-1}x \cap Syy^{-1})\tau_x^{-1} && \text{[by Lemma 5.6(c)]}\\
&= (Sx^{-1}xyy^{-1})\tau_x^{-1} && \text{[by Lemma 5.6(b)]}\\
&= (Sx^{-1}xyy^{-1})\tau_{x^{-1}} && \text{[since } \tau_x^{-1} = \tau_{x^{-1}}\text{]}\\
&= Sx^{-1}xyy^{-1}x^{-1} && \text{[by definition of } \tau_{x^{-1}}\text{]}\\
&= Sxx^{-1}xyy^{-1}x^{-1} && \text{[by Lemma 5.6(c)]}\\
&= Sxyy^{-1}x^{-1}xx^{-1} && \text{[since idempotents commute]}\\
&= Sxyy^{-1}x^{-1}\\
&= S(xy)(xy)^{-1}\\
&= S(xy)^{-1} && \text{[by Lemma 5.6(c)]}\\
&= \operatorname{dom}\tau_{xy}, && \text{[by definition of } \tau_{xy}\text{]}
\end{aligned}$$

and for all $z \in \operatorname{dom}\tau_{xy}$, we have $z\tau_x\tau_y = zxy = z\tau_{xy}$. Hence $(x\varphi)(y\varphi) = \tau_x\tau_y = \tau_{xy} = (xy\varphi)$. Thus φ is a monomorphism. 5.8

Notice that the image of S in $\mathcal{I}_X$ is an inverse subsemigroup of $\mathcal{I}_X$ by Proposition 5.4. However, some subsemigroups of $\mathcal{I}_X$ are not inverse; see Exercise 5.1.

THE NATURAL PARTIAL ORDER

Elements of $\mathcal{I}_X$ are maps, and thus relations, and thus simply subsets of $X\times X$. So we can apply the partial order $\subseteq$ to $\mathcal{I}_X$. However, $\subseteq$ can be characterized using the algebraic structure of $\mathcal{I}_X$, since

$$\begin{aligned}
\sigma \subseteq \tau &\Leftrightarrow \sigma = \tau|_{\operatorname{dom}\sigma}\\
&\Leftrightarrow \sigma = \mathrm{id}_{\operatorname{dom}\sigma}\tau\\
&\Leftrightarrow \sigma = \sigma\sigma^{-1}\tau.
\end{aligned}$$

Since every inverse monoid embeds into $\mathcal{I}_X$ for some X by Theorem 5.8, we can transfer this algebraic definition to arbitrary inverse semigroups by defining $x \preccurlyeq y \Leftrightarrow x = xx^{-1}y$.

Characterizing the relation $\preccurlyeq$

LEMMA 5.9. *For $x, y \in S$, the following are equivalent:*

a) $x \preccurlyeq y$;

b) $x = ey$ *for some* $e \in E(S)$;

c) $x = yf$ *for some* $f \in E(S)$;

d) $x = yx^{-1}x$.

Proof of 5.9. Part 1 [a) $\Rightarrow$ b)]. Suppose $x \preccurlyeq y$. Then $x = xx^{-1}y$, and $e = xx^{-1}$ is an idempotent.

Part 2 [b) $\Rightarrow$ c)]. Suppose $x = ey$. Let $f = y^{-1}ey$. Then

$$f^2 = y^{-1}eyy^{-1}ey = y^{-1}yy^{-1}e^2y = y^{-1}ey = f;$$

thus f is idempotent. Furthermore, $yf = yy^{-1}ey = eyy^{-1}y = ey = x$.

Part 3 [c) $\Rightarrow$ d)]. Suppose $x = yf$. Then $xf = yf^2 = yf = x$ and so $yx^{-1}x = yx^{-1}xf = yfx^{-1}x = xx^{-1}x = x$.

Part 4 [d) $\Rightarrow$ a)]. Suppose $x = yx^{-1}x$. Then $x = yy^{-1}yx^{-1}x = yx^{-1}xy^{-1}y$. Let $e = yx^{-1}xy^{-1}$, so that $x = ey$. Then

$$e^2 = yx^{-1}xy^{-1}yx^{-1}xy^{-1} = yx^{-1}xx^{-1}xy^{-1}yy^{-1} = yx^{-1}xy^{-1} = e,$$

so e is idempotent. Hence $ex = e^2y = ey = x$, and so $exx^{-1} = xx^{-1}$. Thus

$$xx^{-1}y = exx^{-1}y = xx^{-1}ey = xx^{-1}x = x,$$

and so $x \preccurlyeq y$ by definition. 5.9

PROPOSITION 5.10. *The relation $\preccurlyeq$ is a partial order.* $\preccurlyeq$ is a partial order

Proof of 5.10. Since $x = xx^{-1}x$, we have $x \preccurlyeq x$ for any $x \in S$; thus x is reflexive. If $x \preccurlyeq y$ and $y \preccurlyeq x$, then by $x = xx^{-1}y$ and $y = yy^{-1}x$. Hence $x = xx^{-1}yy^{-1}x = yy^{-1}xx^{-1}x = yy^{-1}x = y$; thus $\preccurlyeq$ is anti-symmetric. If $x \preccurlyeq y$ and $y \preccurlyeq z$, then $x = ey$ and $y = fz$ for some $e, f \in E(S)$, and so $x = (ef)z$, and hence $x \preccurlyeq z$ (since ef is in the subsemigroup $E(S)$); thus $\preccurlyeq$ is transitive. 5.10

Proposition 5.10 justifies the choice of the symbol $\preccurlyeq$ for this relation, which is called the *natural partial order* on an inverse semigroup. Notice that if x and y are idempotents, then by the commutativity of idempotents this agrees with the definition of the natural partial order for idempotents (see Proposition 1.19). We are therefore justified in using the same symbol $\preccurlyeq$ for both relations. Natural partial order

PROPOSITION 5.11. a) *The relation $\preccurlyeq$ is compatible (with multiplication); that is, $x \preccurlyeq y \wedge z \preccurlyeq t \Rightarrow xz \preccurlyeq yt$ for all $x, y, z, t \in S$.*

b) *The relation $\preccurlyeq$ is compatible with inversion; that is, $x \preccurlyeq y \Rightarrow x^{-1} \preccurlyeq y^{-1}$ for all $x, y \in S$.*

Proof of 5.11. a) Let $x, y, z, t \in S$. Then

$$\begin{aligned}
&(x \preccurlyeq y) \wedge (z \preccurlyeq t) \\
&\Rightarrow (\exists e, f \in E(S))\big((x = ey) \wedge (z = tf)\big) && \text{[by Lemma 5.9]} \\
&\Rightarrow (\exists e, f \in E(S))\big((xz = eyz) \wedge (yz = ytf)\big) \\
&\Rightarrow (xz \preccurlyeq yz) \wedge (yz \preccurlyeq yt) && \text{[by Lemma 5.9]} \\
&\Rightarrow xz \preccurlyeq yt. && \text{[since } \preccurlyeq \text{ is transitive]}
\end{aligned}$$

b) Let $x, y \in S$. Then

$$\begin{aligned}
& x \preccurlyeq y \\
\Rightarrow\ & (\exists e \in E(S))(x = ey) && \text{[by Lemma 5.9]} \\
\Rightarrow\ & (\exists e \in E(S))(x^{-1} = y^{-1}e) && \text{[by (5.2) and } e^{-1} = e\text{]} \\
\Rightarrow\ & (\exists e \in E(S))(x^{-1} = y^{-1}yy^{-1}e) \\
\Rightarrow\ & (\exists e \in E(S))(x^{-1} = y^{-1}eyy^{-1}) \\
\Rightarrow\ & (\exists f \in E(S))(x^{-1} = fy^{-1}) \\
& && \text{[since } y^{-1}ey \in E(S) \text{ by Lemma 5.6(d)]} \\
\Rightarrow\ & x^{-1} \preccurlyeq y^{-1}. && \text{[by Lemma 5.9]}
\end{aligned}$$

5.11

The natural partial order can serve as a measure of how 'close' an inverse semigroup is to being a group:

Characterizing inverse semigroups that are groups using $\preccurlyeq$

PROPOSITION 5.12. *Let S be an inverse semigroup. Then S is a group if and only if $\preccurlyeq$ is the identity relation on S.*

Proof of 5.12. Suppose S is a group. Then

$$x \preccurlyeq y \Leftrightarrow x = xx^{-1}y \Leftrightarrow x = 1_S y \Leftrightarrow x = y;$$

thus $\preccurlyeq$ is the identity relation.

Now suppose that $\preccurlyeq$ is the identity relation. Let $e, f \in E(S)$. Then $ef \preccurlyeq e$ and $ef \preccurlyeq f$; hence $e = ef = f$. Thus S contains a unique idempotent and so S is a group by Proposition 5.3. 5.12

CLIFFORD SEMIGROUPS

Clifford semigroup

Recall that a semigroup S is a Clifford semigroup if it satisfies the conditions in (4.4). Thus S is a Clifford semigroup if it is completely regular and, for all $x, y \in S$,

$$xx^{-1}yy^{-1} = yy^{-1}xx^{-1}. \tag{5.7}$$

We are going to prove a structure theorem for Clifford semigroups, but first we need to a stronger version of the notion of a semilattice of semigroups, which we introduced in the previous chapter. If we know that S is a semilattice of semigroups S_α, we know something of the coarse structure of S: we know that if $x \in S_\alpha$ and $y \in S_\beta$, then $xy \in S_{\alpha\sqcap\beta}$ (see Figure 4.4).

The new version is stronger in that it describes precisely what products are, rather than simply where they are in the semilattice. Suppose that we

have a semilattice Y, disjoint semigroups S_α for each $\alpha \in Y$, and, for all $\alpha \geqslant \beta$, homomorphisms $\varphi_{\alpha,\beta} : S_\alpha \to S_\beta$ satisfying the conditions

$$(\forall \alpha \in Y)(\varphi_{\alpha,\alpha} = \mathrm{id}_\alpha) \tag{5.8}$$
$$(\forall \alpha, \beta, \gamma \in Y)\big((\alpha \geqslant \beta \geqslant \gamma) \Rightarrow (\varphi_{\alpha,\beta}\varphi_{\beta,\gamma} = \varphi_{\alpha,\gamma})\big) \tag{5.9}$$

Then we can define a multiplication on $S = \bigcup_{\alpha\in Y} S_\alpha$ as follows: for each $x \in S_\alpha$ and $y \in S_\beta$, the product xy is defined to be $(x\varphi_{\alpha,\alpha\sqcap\beta})(y\varphi_{\beta,\alpha\sqcap\beta})$. That is, we use the homomorphisms to map x and y 'down' into $S_{\alpha\sqcap\beta}$ and multiply them there; see Figure 5.2. For any $x \in S_\alpha$, $y \in S_\beta$, $z \in S_\gamma$,

$$\begin{aligned}
&x(yz)\\
&= x\big((y\varphi_{\beta,\beta\sqcap\gamma})(z\varphi_{\gamma,\beta\sqcap\gamma})\big) && \text{[by definition of multiplication]}\\
&= (x\varphi_{\alpha,\alpha\sqcap\beta\sqcap\gamma})\big((y\varphi_{\beta,\beta\sqcap\gamma})(z\varphi_{\gamma,\beta\sqcap\gamma})\big)\varphi_{\beta\sqcap\gamma,\alpha\sqcap\beta\sqcap\gamma}\\
&&& \text{[by definition of multiplication]}\\
&= (x\varphi_{\alpha,\alpha\sqcap\beta\sqcap\gamma})(y\varphi_{\beta,\beta\sqcap\gamma}\varphi_{\beta\sqcap\gamma,\alpha\sqcap\beta\sqcap\gamma})(z\varphi_{\gamma,\beta\sqcap\gamma}\varphi_{\beta\sqcap\gamma,\alpha\sqcap\beta\sqcap\gamma})\\
&&& \text{[since } \varphi_{\beta\sqcap\gamma,\alpha\sqcap\beta\sqcap\gamma} \text{ is a homomorphism]}\\
&= (x\varphi_{\alpha,\alpha\sqcap\beta\sqcap\gamma})\big((y\varphi_{\beta,\alpha\sqcap\beta\sqcap\gamma})(z\varphi_{\gamma,\alpha\sqcap\beta\sqcap\gamma})\big) && \text{[by (5.9)]}\\
&= \big((x\varphi_{\alpha,\alpha\sqcap\beta\sqcap\gamma})(y\varphi_{\beta,\alpha\sqcap\beta\sqcap\gamma})\big)(z\varphi_{\gamma,\alpha\sqcap\beta\sqcap\gamma}) && \text{[by associativity in } S_{\alpha\sqcap\beta\sqcap\gamma}]\\
&= (xy)z, && \text{[by similar reasoning]}
\end{aligned}$$

and so this multiplication is associative. This semigroup S is a *strong semilattice of semigroups* and is denoted $\mathcal{S}[Y; S_\alpha; \varphi_{\alpha,\beta}]$. If every S_α is a group, it is a *strong semilattice of groups*.

Strong semilattice of semigroups/groups

An element x of a semigroup S is *central* if $xy = yx$ for all $y \in S$.

Central element

THEOREM 5.13. *The following are equivalent:*

a) *S is a Clifford semigroup;*
b) *S is a semilattice of groups;*
c) *S is a strong semilattice of groups;*
d) *S is regular, and the idempotents of S are central;*
e) *S is regular, and every $\mathcal{D}$-class of S contains a unique idempotent.*

Characterization of Clifford semigroups

Proof of 5.13. Part 1 [a) $\Rightarrow$ b)]. Let S be a Clifford semigroup. Then S is completely regular and so is a semilattice of completely simple semigroups S_α by Theorem 4.17. Let e, f be idempotents. Then $e = ee^{-1}e = eee^{-1} = ee^{-1}$ by (4.2) and similarly $f = ff^{-1}$ and so $ef = fe$ by (5.7). So all idempotents of S commute. Now, S_α is completely simple and so $S_\alpha \simeq \mathcal{M}[G; I, \Lambda; P]$ for some group G, index sets I and Λ, and matrix P over G. Let $e, f \in S_\alpha$ be idempotents. Then $e = (i, p_{\lambda i}^{-1}, \lambda)$ and $f = (j, p_{\mu j}^{-1}, \mu)$, and $(i, p_{\lambda i}^{-1} p_{\lambda j} p_{\mu j}^{-1}, \mu) = ef = fe = (j, p_{\mu j}^{-1} p_{\mu i} p_{\lambda i}^{-1}, \lambda)$. Hence $i = j$ and $\lambda = \mu$ and so $e = f$. So each S_α contains only one idempotent. Thus, by Proposition 4.14, S_α is a group. Therefore S is a semilattice of groups.

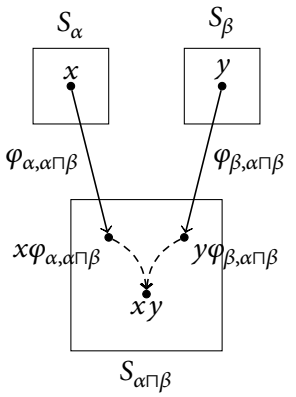


FIGURE 5.2
Multiplying in a strong semilattice of semigroups

Part 2 [b) ⇒ c)]. Let S be a semilattice of groups S_α, where $\alpha \in Y$. To prove that S is a *strong* semilattice of groups, we have to define homomorphisms $\varphi_{\alpha,\beta}$ for all $\alpha, \beta \in Y$, prove that (5.8) and (5.9) hold, and show that the strong semilattice of groups $\mathcal{S}[Y; S_\alpha; \varphi_{\alpha,\beta}]$ is isomorphic to S.

Write 1_α for the identity of the group S_α. Then for $\alpha \geqslant \beta$ and $x \in S_\alpha$, we have $1_\beta x \in S_\beta$. Hence we can define a map $\varphi_{\alpha,\beta} : S_\alpha \to S_\beta$ by $x\varphi_{\alpha,\beta} = 1_\beta x$. For $x, y \in S_\alpha$,

$$\begin{aligned}
&(x\varphi_{\alpha,\beta})(y\varphi_{\alpha,\beta}) \\
&= 1_\beta x 1_\beta y && \text{[by definition of } \varphi_{\alpha,\beta}\text{]} \\
&= 1_\beta xy && \text{[since } 1_\beta x \in S_\beta \text{ and thus } (1_\beta x)1_\beta = 1_\beta x\text{]} \\
&= (xy)\varphi_{\alpha,\beta}; && \text{[by definition of } \varphi_{\alpha,\beta}\text{]}
\end{aligned}$$

hence $\varphi_{\alpha,\beta}$ is a homomorphism. Clearly $\varphi_{\alpha,\alpha} = \mathrm{id}_{S_\alpha}$, so (5.8) holds. For $\alpha \geqslant \beta \geqslant \gamma$, for any $x \in S_\alpha$

$$\begin{aligned}
&x\varphi_{\alpha,\beta}\varphi_{\beta,\gamma} \\
&= (1_\beta x)\varphi_{\beta,\gamma} && \text{[by definition of } \varphi_{\alpha,\beta}\text{]} \\
&= 1_\gamma 1_\beta x && \text{[by definition of } \varphi_{\beta,\gamma}\text{]} \\
&= (1_\beta \varphi_{\beta,\gamma})x && \text{[by definition of } \varphi_{\beta,\gamma}\text{]} \\
&= 1_\gamma x && \text{[by Corollary 5.5]} \\
&= x\varphi_{\alpha,\gamma}; && \text{[by definition of } \varphi_{\alpha,\gamma}\text{]}
\end{aligned}$$

hence (5.9) holds. Finally, for any $x \in S_\alpha$ and $y \in S_\beta$,

$$\begin{aligned}
xy &= 1_{\alpha\sqcap\beta} xy && \text{[since } xy \in S_{\alpha\sqcap\beta}\text{]} \\
&= 1_{\alpha\sqcap\beta} x 1_{\alpha\sqcap\beta} y && \text{[since } 1_{\alpha\sqcap\beta} x \in S_{\alpha\sqcap\beta}\text{]} \\
&= (x\varphi_{\alpha,\alpha\sqcap\beta})(y\varphi_{\beta,\alpha\sqcap\beta}). && \text{[by definition of } \varphi_{\alpha,\alpha\sqcap\beta} \text{ and } \varphi_{\alpha,\alpha\sqcap\beta}\text{]}
\end{aligned}$$

Therefore S is isomorphic to $\mathcal{S}[Y; S_\alpha; \varphi_{\alpha,\beta}]$.

Part 3 [c) ⇒ d)]. A strong semilattice of groups $S = \mathcal{S}[Y; S_\alpha; \varphi_{\alpha,\beta}]$ is certainly regular: for each $x \in S_\alpha$, let x^{-1} be the inverse of x in the group S_α. The idempotents of S are the identities of the groups S_α. Write 1_α for the identity of S_α. Then for any $\beta \in Y$ and $x \in S_\beta$,

$$\begin{aligned}
1_\alpha x &= (1_\alpha \varphi_{\alpha,\alpha\sqcap\beta})(x\varphi_{\beta,\alpha\sqcap\beta}) = 1_{\alpha\sqcap\beta}(x\varphi_{\beta,\alpha\sqcap\beta}) \\
&= (x\varphi_{\beta,\alpha\sqcap\beta}) = (x\varphi_{\beta,\alpha\sqcap\beta})1_{\alpha\sqcap\beta} = (x\varphi_{\beta,\alpha\sqcap\beta})(1_\alpha\varphi_{\alpha,\alpha\sqcap\beta}) = x1_\alpha.
\end{aligned}$$

Thus every idempotent of S is central.

Part 4 [d) ⇒ e)]. Each $\mathcal{D}$-class D_x must contain at least one idempotent, namely xx^{-1}. Suppose e and f are idempotent and $e \mathrel{\mathcal{D}} f$. Then by

Proposition 3.21(b) there exists an element x and inverse x' such that $xx' = e$ and $x'x = f$. Therefore

$$\begin{aligned}
e &= e^2 \\
&= xx'xx' && [\text{since } xx' = e] \\
&= xfx' && [\text{since } x'x = f] \\
&= xx'f && [\text{since } f \text{ is central}] \\
&= xx'x'x && [\text{since } f = x'x] \\
&= ex'x && [\text{since } xx' = e] \\
&= x'ex && [\text{since } e \text{ is central}] \\
&= x'xx'x && [\text{since } e = xx'] \\
&= f^2 = f. && [\text{since } f = x'x]
\end{aligned}$$

Hence every $\mathcal{D}$-class of S contains a unique idempotent.

Part 5 [e) $\Rightarrow$ a)]. Since every $\mathcal{D}$-class contains a unique idempotent, every $\mathcal{D}$-class consists of a single $\mathcal{H}$-class by Proposition 3.20, and so $\mathcal{D} = \mathcal{H}$. Furthermore, each of these $\mathcal{H}$-classes is a group by Proposition 3.14, and so every element of S lies in a subgroup and thus S is completely regular by Theorem 4.15. Thus, by Theorem 4.17, S is a semilattice of completely simple semigroups S_α. Every element of a completely simple semigroup is $\mathcal{D}$-related, and so every S_α is contained within a single $\mathcal{D}$-class and is thus a group. So S is a semilattice of groups and thus, by the second part of this proof, a strong semilattice of groups $\mathcal{S}[Y; S_\alpha; \varphi_{\alpha,\beta}]$. Hence for $x \in S_\alpha$ and $y \in S_\beta$, we have $xx^{-1}yy^{-1} = 1_\alpha 1_\beta = 1_{\alpha\sqcap\beta} = 1_\beta 1_\alpha = yy^{-1}xx^{-1}$. 5.13

In particular, Theorem 5.13(d) implies that in a Clifford semigroup, idempotents commute; hence, by Theorem 5.1, Clifford semigroups are inverse semigroups. Notice that this is not obvious from the conditions (4.3) and (4.4).

Natural partial order on Clifford semigroups

Let S be a Clifford semigroup. By Theorem 5.13, S is isomorphic to a strong semilattice of groups $\mathcal{S}[Y; G_\alpha; \varphi_{\alpha,\beta}]$. Let $x \in G_\alpha$ and $y \in G_\beta$. Then

$$\begin{aligned}
x \preccurlyeq y &\Leftrightarrow x = (xx^{-1})y \\
&\Leftrightarrow x = 1_\alpha y \\
&\Leftrightarrow x = (1_\alpha \varphi_{\alpha,\alpha\sqcap\beta})(y\varphi_{\beta,\alpha\sqcap\beta}) \\
&\Leftrightarrow \big(x = (1_\alpha\varphi_{\alpha,\alpha})(y\varphi_{\beta,\alpha})\big) \wedge (\alpha \sqcap \beta = \alpha) \\
&\Leftrightarrow (x = y\varphi_{\beta,\alpha\sqcap\beta}) \wedge (\alpha \leqslant \beta).
\end{aligned}$$

Thus the natural partial order $\preccurlyeq$ precisely corresponds to the homomorphisms $\varphi_{\alpha,\beta}$ and the order of the semilattice $(Y, \leqslant)$. In particular, we have

$$1_\alpha \preccurlyeq 1_\beta \Leftrightarrow 1_\alpha = 1_\beta\varphi_{\beta,\alpha} \wedge (\alpha \leqslant \beta) \Leftrightarrow \alpha \leqslant \beta.$$

Since the identities of the groups G_α are precisely the idempotents of S, we see that $(E(S), \preccurlyeq)$ and $(Y, \leqslant)$ are isomorphic. In particular, every

semilattice $(Y, \leqslant)$ is a Clifford semigroup $\mathcal{S}[Y; G_\alpha, \varphi_{\alpha,\beta}]$ where the groups G_α are all trivial.

FREE INVERSE SEMIGROUPS

Let A be an alphabet. Let A^{-1} be a set of new symbols bijection with A under the map $a \mapsto a^{-1}$. Extend this map to an involution of $A \cup A^{-1}$ by defining $(a^{-1})^{-1} = a$. For any word $a_1 a_2 \cdots a_n \in (A \cup A^{-1})^*$, define $(a_1 a_2 \cdots a_n)^{-1} = a_n^{-1} \cdots a_2^{-1} a_1^{-1}$. Let $\mathrm{FInvS}(A)$ be semigroup presented by $\mathrm{Sg}\langle A \cup A^{-1} \mid \rho \rangle$, where

$$\begin{aligned}\rho = {} & \{ (uu^{-1}u, u) : u \in (A \cup A^{-1})^+ \} \\ & \cup \{ (uu^{-1}vv^{-1}, vv^{-1}uu^{-1}) : u, v \in (A \cup A^{-1})^+ \}.\end{aligned}$$

PROPOSITION 5.14. *The semigroup* $\mathrm{FInvS}(A)$ *is an inverse semigroup, where the inverse of* $[u]_{\rho^\#} \in \mathrm{FInvS}(A)$ *is* $[u^{-1}]_{\rho^\#}$.

Proof of 5.14. Define $\left([u]_{\rho^\#}\right)^{-1} = [u^{-1}]_{\rho^\#}$. We aim to prove that the conditions (5.1)–(5.4) are satisfied. First of all, it is necessary to check that the operation $^{-1}$ is well-defined on $\mathrm{FInvS}(A)$. Suppose $[u]_{\rho^\#} = [v]_{\rho^\#}$. Then there is a sequence of elementary ρ-transitions $u = w_0 \leftrightarrow_\rho w_1 \leftrightarrow_\rho \ldots \leftrightarrow_\rho w_n = v$. Apply $^{-1}$ (as an operation on $(A \cup A^{-1})^+$) to every term in this sequence. This yields a sequence of elementary ρ-transitions $u^{-1} = w_0^{-1} \leftrightarrow_\rho w_1^{-1} \leftrightarrow_\rho \ldots \leftrightarrow_\rho w_n^{-1} = v^{-1}$; hence $[u^{-1}]_{\rho^\#} = [v^{-1}]_{\rho^\#}$.

Now let $u, v \in \mathrm{FInvS}(A)$. It is immediate from the definition of $^{-1}$ that

$$([u]_{\rho^\#}^{-1})^{-1} = [(u^{-1})^{-1}]_{\rho^\#} = [u]_{\rho^\#}$$

and

$$\begin{aligned}[uv]_{\rho^\#}^{-1} = [(uv)^{-1}]_{\rho^\#} &= [v^{-1}u^{-1}]_{\rho^\#} \\ &= [v^{-1}]_{\rho^\#}[u^{-1}]_{\rho^\#} = [v]_{\rho^\#}^{-1}[u]_{\rho^\#}^{-1};\end{aligned}$$

thus (5.1) and (5.2) hold. Furthermore,

$$\begin{aligned}[u]_{\rho^\#}[u]_{\rho^\#}^{-1}[u]_{\rho^\#} &= [u]_{\rho^\#}[u^{-1}]_{\rho^\#}[u]_{\rho^\#} \\ &= [uu^{-1}u]_{\rho^\#} && \text{[by definition of } \rho\text{]} \\ &= [u]_{\rho^\#}\end{aligned}$$

and

$$\begin{aligned}[u]_{\rho^\#}[u]_{\rho^\#}^{-1}[v]_{\rho^\#}[v]_{\rho^\#}^{-1} &= [u]_{\rho^\#}[u^{-1}]_{\rho^\#}[v]_{\rho^\#}[v^{-1}]_{\rho^\#} \\ &= [uu^{-1}vv^{-1}]_{\rho^\#}\end{aligned}$$

$$
\begin{aligned}
&= [vv^{-1}uu^{-1}]_{\rho^\#} && \text{[by definition of } \rho\text{]}\\
&= [v]_{\rho^\#}[v^{-1}]_{\rho^\#}[u]_{\rho^\#}[u^{-1}]_{\rho^\#}\\
&= [v]_{\rho^\#}[v]_{\rho^\#}^{-1}[u]_{\rho^\#}[u]_{\rho^\#}^{-1};
\end{aligned}
$$

thus (5.3) and (5.4) hold. Hence $\mathrm{FInvS}(A)$ is an inverse semigroup. 5.14

Free inverse semigroup

Let F be an inverse semigroup, let A be an alphabet, and let $\iota : A \to F$ be an embedding of A into F. Then the inverse semigroup F is a *free inverse semigroup on A* if, for any inverse semigroup S and map $\varphi : A \to S$, there is a unique homomorphism $\overline{\varphi} : F \to S$ that extends φ (that is, with $\iota\overline{\varphi} = \varphi$). Using diagrams, this definition says that F is a free inverse semigroup on A if

$$
\left.\begin{array}{l}
\text{for all } \begin{array}{ccc} A & \xhookrightarrow{\iota} & F \\ & {\scriptstyle\varphi}\searrow & \\ & & S \end{array} \text{ with } S \text{ inverse, there exists}\\[2ex]
\text{a unique homomorphism } \overline{\varphi} \text{ such that } \begin{array}{ccc} A & \xhookrightarrow{\iota} & F \\ & {\scriptstyle\varphi}\searrow & \downarrow{\scriptstyle\overline{\varphi}} \\ & & S \end{array}.
\end{array}\right\} \quad (5.10)
$$

This definition is analogous to the definition of the free semigroup on A (see pages 38 f.). In Chapter 8, we will see definitions of 'free objects' in a much more general setting. Like the free semigroup on A, the free inverse semigroup on A is unique up to isomorphism:

Uniqueness of the free inverse semigroup on A

PROPOSITION 5.15. *Let A be an alphabet and let F be an inverse semigroup. Then F is a free inverse semigroup on A if and only if $F \simeq \mathrm{FInvS}(A)$.*

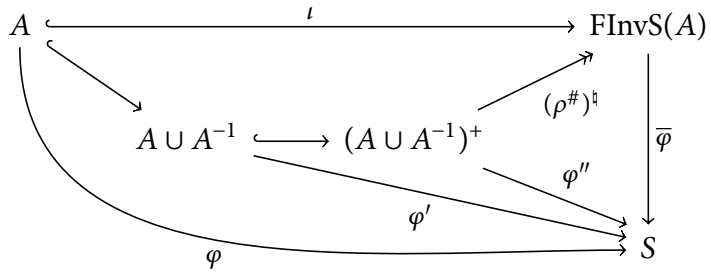
Proof of 5.15. Let $\iota : A \to \mathrm{FInvS}(A)$ be the natural map $a\iota = [a]_{\rho^\#}$. Let S be a inverse semigroup and $\varphi : A \to S$ a map. Extend φ to a map $\varphi' : A \cup A^{-1} \to S$ by defining $a^{-1}\varphi' = (a\varphi)^{-1}$ for $a^{-1} \in A^{-1}$. Since $(A \cup A^{-1})^+$ is the free semigroup on A, the map φ' extends to a unique homomorphism $\varphi'' : (A \cup A^{-1})^+ \to S$ with $(a_1a_2\cdots a_n)\varphi'' = (a_1\varphi')(a_2\varphi')\cdots(a_n\varphi')$, where $a_i \in A \cup A^{-1}$. Since S is an inverse semigroup,

$$
\begin{aligned}
(uu^{-1}u)\varphi'' &= (u\varphi'')(u^{-1}\varphi'')(u\varphi'') && \text{[since } \varphi'' \text{ is a homomorphism]}\\
&= (u\varphi'')(u\varphi'')^{-1}(u\varphi'') && \text{[by definition of } \varphi'\text{]}\\
&= u\varphi''
\end{aligned}
$$

and

$$
\begin{aligned}
(uu^{-1}vv^{-1})\varphi'' &= (u\varphi'')(u^{-1}\varphi'')(v\varphi'')(v^{-1}\varphi'')\\
&= (u\varphi'')(u\varphi'')^{-1}(v\varphi'')(v\varphi'')^{-1}\\
&= (v\varphi'')(v\varphi'')^{-1}(u\varphi'')(u\varphi'')^{-1} && \text{[since } S \text{ is inverse]}\\
&= (uu^{-1}vv^{-1})\varphi''
\end{aligned}
$$

for all $u, v \in (A \cup A^{-1})^+$. Thus $\rho \subseteq \ker \varphi''$ and so there is a well-defined homomorphism $\overline{\varphi} : \mathrm{FInvS}(A) \to S$ with $[u]_{\rho^\#}\overline{\varphi} = u\varphi''$. That is, the following diagram commutes:



It remains to prove that $\overline{\varphi}$ is the unique homomorphism such that $\iota\overline{\varphi} = \varphi$. So let $\psi : \mathrm{FInvS}(A) \to S$ be such that $\iota\psi = \varphi$. Then for all $a \in A$, we have $[a]_{\rho^\#}\psi = a\iota\psi = a\varphi$ and

$$\begin{aligned}
[a^{-1}]_{\rho^\#}\psi &= ([a]_{\rho^\#})^{-1}\psi && \text{[by definition of } {}^{-1} \text{ in FInvS}(A)\text{]}\\
&= ([a]_{\rho^\#}\psi)^{-1} && \text{[by Proposition 5.4]}\\
&= (a\iota\psi)^{-1}\\
&= (a\varphi)^{-1}\\
&= a^{-1}\varphi. && \text{[by Proposition 5.4]}
\end{aligned}$$

Hence for any $a_i \in A \cup A^{-1}$,

$$\begin{aligned}
([a_1a_2\cdots a_n]_{\rho^\#})\psi &= ([a_1]_{\rho^\#}[a_2]_{\rho^\#}\cdots[a_n]_{\rho^\#})\psi\\
&= ([a_1]_{\rho^\#}\psi)([a_2]_{\rho^\#}\psi)\cdots([a_n]_{\rho^\#})\psi)\\
&= (a_1\varphi)(a_2\varphi)\cdots(a_n\varphi)\\
&= ([a_1]_{\rho^\#}\varphi)([a_2]_{\rho^\#}\varphi)\cdots([a_n]_{\rho^\#})\varphi)\\
&= ([a_1a_2\cdots a_n]_{\rho^\#})\overline{\varphi}.
\end{aligned}$$

Thus $\psi = \overline{\varphi}$. Therefore $\mathrm{FInvS}(A)$ is a free inverse semigroup on A.

Now let F be a free inverse semigroup on A. Let $\iota_1 : A \to \mathrm{FInvS}(A)$ and $\iota_2 : A \to F$ be the embedding maps. Following the same argument as for free semigroups on A (see the proof of Proposition 2.1), this leads to $\bar{\iota}_2 : \mathrm{FInvS}(A) \to F$ and $\bar{\iota}_1 : F \to \mathrm{FInvS}(A)$ being mutually inverse isomorphisms. 5.15

Free inverse monoid

We could repeat the discussion above for monoids instead of semigroups. The monoid $\mathrm{FInvM}(A)$ is presented by $\mathrm{Mon}\langle A \cup A^{-1} \mid \rho\rangle$. A monoid F is a *free inverse monoid on* A if, for any inverse monoid S and map $\varphi : A \to S$, there is a unique monoid homomorphism $\overline{\varphi} : F \to S$ extending φ; that is, with $\iota\overline{\varphi} = \varphi$. One can prove an analogy of Proposition 5.15 for monoids, showing that an inverse monoid F is a free inverse monoid on A if and only if $F \simeq \mathrm{FInvM}(A)$. Notice that because there is

no defining relation in ρ that has the empty word ε as one of its two sides, there is no non-empty word that is equal to ε in $\mathrm{FInvM}(A)$. Therefore $\mathrm{FInvM}(A) \simeq (\mathrm{FInvS}(A))^1$.

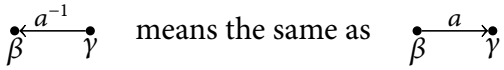
Since free inverse semigroups and monoids are such fundamental objects, we would like to be able to solve the word problem: given two words in $(A \cup A^{-1})^+$ (respectively, $(A \cup A^{-1})^*$), do they represent the same element of $\mathrm{FInvS}(A)$ (respectively, $\mathrm{FInvM}(A)$)? This appears difficult: for example,

$$\left.\begin{aligned} &aaa^{-1}a^{-1}a^{-1}abb^{-1}ab^{-1}bcaa^{-1}cc^{-1} \\ &\quad =_{\mathrm{FInvS}(A)} a^{-1}abb^{-1}aaa^{-1}caa^{-1}cc^{-1}c^{-1}b^{-1}ba^{-1}ac, \end{aligned}\right\} \quad (5.11)$$

but this not obvious. However, we now introduce a representation of elements of $\mathrm{FInvM}(A)$ that makes it easy to answer this question.

Tree with edge labels from $A \cup A^{-1}$

Let T be a finite non-empty directed tree with edges labelled by symbols in A. Extend the set of labels to $A \cup A^{-1}$ by adopting the following convention: for all $a \in A$ and vertices β and γ,



(5.12)

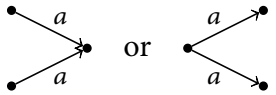
Denote the set of vertices of T by $V(T)$. By definition, $|V(T)| \geqslant 1$. Let $\beta, \gamma \in V(T)$. If β and γ are adjacent, then $\beta\gamma$ will denote the edge from β to γ. A (β, γ)-*walk* on T is a sequence $\beta = \delta_0, \delta_1, \ldots, \delta_n = \gamma$ such that δ_{i-1} and δ_i are adjacent for $i = 1, \ldots, n$. A (β, γ)-walk $\beta = \delta_0, \delta_1, \ldots, \delta_n = \gamma$ *spans* T if every vertex of T appears at least once among the δ_i. The (β, γ)-*path* on T, denoted $\pi(\beta, \gamma)$, is the unique (β, γ)-walk $\beta = \delta_0, \delta_1, \ldots, \delta_n = \gamma$ such that no vertex of T occurs more than once among the δ_i; the integer n is the *length* of $\pi(\beta, \gamma)$. Notice that there is a *trivial path* at β, namely $\pi(\beta, \beta)$, which has length 0.

For a (β, γ)-walk $\sigma = (\beta = \delta_0, \ldots, \delta_m = \gamma)$, define $\mathrm{w}(\sigma) = x_1 x_2 \cdots x_m$, where $x_i \in A \cup A^{-1}$ is the label on the edge $\delta_{i-1}\delta_i$ for $i = 1, \ldots, m$ (recalling the convention (5.12)). Note that $\mathrm{w}(\pi(\beta, \beta)) = \varepsilon$.

Word tree, Munn tree

A *word tree* over A is a finite non-empty directed tree T with edges labelled elements of A (using the convention (5.12)), and where there is no vertex that has two distinct incoming edges with the same label or two distinct outgoing edges with the same label. That is,

a word tree does not contain subgraphs



(5.13)

A *Munn tree* over A is a word tree T with two distinguished vertices α_T and ω_T (not necessarily distinct).

Figure 5.3 gives an example of a Munn tree. Notice it satisfies the condition (5.13). Furthermore, both words in (5.11) label spanning (α_T, ω_T)-walks in this Munn tree. This is how Munn trees allow us to solve the

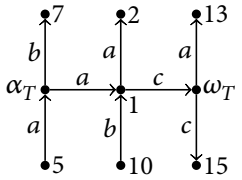


FIGURE 5.3
Munn tree T for the words $a^2a^{-3}abb^{-1}ab^{-1}bcaa^{-1}cc^{-1}$ and $a^{-1}abb^{-1}a^2a^{-1}caa^{-1}cc^{-2}b^{-1}ba^{-1}ac$.

word problem for FInvM(A): we will prove that two words represent the same element of FInvM(A) if and only if they have isomorphic Munn trees.

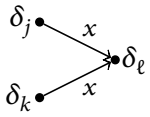
To be precise, an isomorphism between two word trees T_1 and T_2 is a bijection $\varphi : V(T_1) \to V(T_2)$ such that there is an edge vv' labelled by a in T_1 if and only if there is an edge $(v\varphi)(v'\varphi)$ labelled by a in T_2. If T_1 and T_2 are Munn trees, then such a map is an isomorphism if, in addition, $\alpha_{T_1}\varphi = \alpha_{T_2}$ and $\omega_{T_1}\varphi = \omega_{T_2}$.

Constructing a Munn tree from a word

Suppose we have a word $u = x_1 x_2 \cdots x_n$, where $x_i \in A \cup A^{-1}$. Let us describe how to construct a Munn tree T with a spanning (α_T, ω_T)-walk σ such that $\mathrm{w}(\sigma) = u$. We will initially construct a tree T with distinguished vertices α_T and ω_T such that there is a spanning (α_T, ω_T)-walk σ on T such that $\mathrm{w}(\sigma) = u$. This tree may not satisfy (5.13). We will then modify T to turn it into a Munn tree.

To begin, let T be the graph with $n + 1$ vertices $\delta_0, \delta_1, \delta_2, \ldots, \delta_{n-1}, \delta_n$ with edges $\delta_{i-1}\delta_i$ having label x_i for $i = 1, \ldots, n$ (recall the convention (5.12)). Notice that this tree is simply a path. Let $\alpha_T = \delta_0$ and $\omega_T = \delta_n$. Note that T is a tree with distinguished vertices α_T and ω_T. Let σ be unique path from α_T to ω_T. Then $\mathrm{w}(\sigma) = u$. (The graph at the top of Figure 5.4 is the result of this construction for $u = a^2a^{-3}abb^{-1}ab^{-1}bcaa^{-1}cc^{-1}$.) Note that T satisfies all the conditions we want except possibly (5.13). Now let us modify T.

If T satisfies (5.13), then it is a word tree and so a Munn tree and we are finished. So suppose T does not satisfy (5.13). Then by the convention (5.12), T contains a subgraph



for some $x \in A \cup A^{-1}$.

Fix such a subgraph. Modify T by folding the (identically-labelled) edges $\delta_j\delta_\ell$ and $\delta_j\delta_\ell$ together and merging the vertices δ_j and δ_k. If we merge α_T (respectively, ω_T) with some vertex, the resulting merged vertex is still α_T (respectively, ω_T). Then T is still a tree and the walk σ (which is, after all, simply a sequence of vertices) is still a spanning (α_T, ω_T)-walk for T. However, T now contains one vertex fewer than before.

Repeat this process. Since each such modification reduces the number of vertices of T, then process must halt with a tree T satisfying (5.13), which is the desired Munn tree. (Figure 5.4 illustrates this process for the word $u = a^2a^{-3}abb^{-1}ab^{-1}bcaa^{-1}cc^{-1}$.)

We now establish four lemmata that lead up to the main result. For brevity, we write F for FInvM(A). First, we must make some more definitions.

Let $\sigma = (\beta = \delta_0, \ldots, \delta_m = \gamma)$ and $\tau = (\gamma = \eta_0, \ldots, \eta_n = \zeta)$ be,

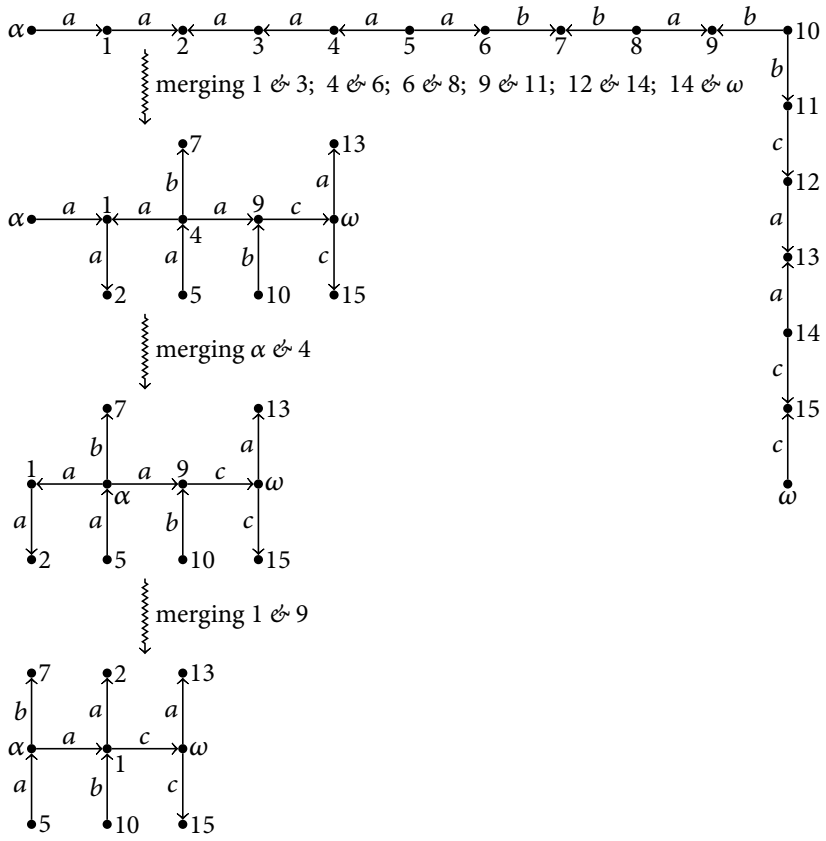


FIGURE 5.4
Folding a linear graph to produce a Munn tree for the word $a^2a^{-3}abb^{-1}ab^{-1}bcaa^{-1}cc^{-1}$.

respectively, a (β, γ)- and a (γ, ζ)-walk on T. Define a (β, ζ)-walk $\sigma\tau$ by

$$\sigma\tau = (\beta = \delta_0, \dots, \delta_{m-1}, \gamma, \eta_1, \dots, \eta_n = \zeta).$$

Clearly one can extend this to products of three or more walks and this product is associative (whenever it is defined). We also define σ^{-1} to be the (γ, β)-walk $(\gamma = \delta_m, \dots, \delta_0 = \beta)$. If σ is a (β, β)-walk, then σ^k has the obvious meaning for all $k \in \mathbb{N}$.

LEMMA 5.16. *Let σ be a (β, γ)-walk and τ a (γ, ζ)-walk on a word tree T. Then:*

a) $\mathrm{w}(\sigma\tau) = \mathrm{w}(\sigma)\,\mathrm{w}(\tau)$;

b) $\tau = \sigma^{-1}$ *if and only if* $\mathrm{w}(\tau) = (\mathrm{w}(\sigma))^{-1}$.

Proof of 5.16. Part a) and the forward implication in part b) are immediate from the definition. It remains to prove the reverse implication in part b). So suppose $\mathrm{w}(\tau) = (\mathrm{w}(\sigma))^{-1} = x_1 \cdots x_m$ (where $x_i \in A \cup A^{-1}$). Then σ and τ both contain $m + 1$ vertices, with $\sigma = (\beta = \delta_0, \dots, \delta_m = \gamma)$ and $\tau = (\gamma = \eta_0, \dots, \eta_m = \zeta)$. We will prove that $\delta_{m-j} = \eta_j$ by induction on j. We already know that $\delta_m = \gamma = \eta_0$; this is the base of the induction. For the induction step, suppose that $\delta_{m-j} = \eta_j$. Now, $\delta_{m-j}\delta_{m-j-1}$ and $\eta_j\eta_{j+1}$ both have label x_j. So, since T satisfies (5.13), $\delta_{m-j-1} = \eta_{j+1}$. This proves the induction step. So $\delta_{m-j} = \eta_j$ for all $j = 1, \dots, m$. Hence $\tau = \sigma^{-1}$. 5.16

The next lemma essentially says that each Munn tree is associated with a unique element of $\mathrm{FInvM}(A)$:

LEMMA 5.17. *If σ and τ are spanning (α_T, ω_T)-walks on a Munn tree T, then* $\mathrm{w}(\sigma) =_F \mathrm{w}(\tau)$.

Proof of 5.17. If $|V(T)| = 1$, then σ and τ consist of the single vertex in $V(T)$ and so $\mathrm{w}(\sigma) = \varepsilon = \mathrm{w}(\tau)$.

In the remaining cases, we use induction on $|V(T)| \geqslant 2$ to prove the following statement: If σ and τ are spanning (β, γ)-walks on a word tree T, then $\mathrm{w}(\sigma) =_F \mathrm{w}(\tau)$.

For the base of the induction, let $|V(T)| = 2$. Let ζ be the unique vertex in $T \setminus \{\beta\}$, let $\pi = \pi(\beta, \zeta)$ and let $x = \mathrm{w}(\pi)$. Note that $x \in A \cup A^{-1}$ since π has length 1, because there are only two vertices in T. We now consider the cases $\gamma = \beta$ and $\gamma = \zeta$ separately:

- $\gamma = \beta$. Then $\sigma = (\pi\pi^{-1})^k$ and $\tau = (\pi\pi^{-1})^\ell$ for some $k, \ell \in \mathbb{N} \cup \{0\}$ and so, by Lemma 5.16 and the defining relations in ρ,

$$\mathrm{w}(\sigma) = (xx^{-1})^k =_F xx^{-1} =_F (xx^{-1})^\ell = \mathrm{w}(\tau).$$

- $\gamma = \zeta$. Then $\sigma = (\pi\pi^{-1})^k\pi$ and $\tau = (\pi\pi^{-1})^\ell\pi$ for some $k, \ell \in \mathbb{N} \cup \{0\}$ and so, by Lemma 5.16 and the defining relations in ρ,

$$\mathrm{w}(\sigma) = (xx^{-1})^k x =_F x =_F (xx^{-1})^\ell x = \mathrm{w}(\tau).$$

In either case, the result holds for $|V(T)| = 2$.

For the inductive step, let $n > 2$. Suppose that if $\overline{\sigma}$ and $\overline{\tau}$ are spanning (β, γ) walks on a tree $\overline{T}$ such that $|V(\overline{T})| < n$, then $\mathrm{w}(\overline{\sigma}) =_F \mathrm{w}(\overline{\tau})$.

CLAIM. If σ_0 is a (ξ, ξ)-walk on a subtree $\overline{T}$ of T such that $|V(\overline{T})| < n$, then $(\mathrm{w}(\sigma_0))^2 =_F \mathrm{w}(\sigma_0)$.

Proof of Claim. Let $\overline{T}_0$ be the subtree of $\overline{T}$ spanned by σ_0. Then we have $|V(\overline{T}_0)| \leqslant |V(\overline{T})| < n$ and both σ_0 and σ_0^2 are spanning (ξ, ξ)-walks on $\overline{T}_0$. Thus, by the induction hypothesis, $\mathrm{w}(\sigma_0) =_F \mathrm{w}(\sigma_0^2) = (\mathrm{w}(\sigma_0))^2$. Claim

Now let σ and τ be spanning (β, γ)-walks on T. We consider separately the case where β is an endpoint of T and the case where β is not an endpoint of T.

- β is an endpoint (or leaf vertex) of T. Let ξ be the unique vertex of T adjacent to β and let $\overline{T}$ be the subtree of T obtained by deleting β and the edge $\beta\xi$. Let $\pi = \pi(\beta, \xi)$ and let $x = \mathrm{w}(\pi)$. Now we consider the sub-cases $\beta = \gamma$ and $\beta \neq \gamma$ separately.
  - $\gamma = \beta$. Then for some (ξ, ξ)-walks $\sigma_1, \sigma_2, \ldots, \sigma_h$ on $\overline{T}$ and some $k_i \in \mathbb{N} \cup \{0\}$ (where $i = 0, \ldots, h$),

$$\sigma = \pi(\pi^{-1}\pi)^{k_0}\sigma_1(\pi^{-1}\pi)^{k_1}\sigma_2 \cdots \sigma_h(\pi^{-1}\pi)^{k_h}\pi^{-1}.$$

Let $u_i = \mathrm{w}(\sigma_i)$ for $i = 1, \dots, h$. By Lemma 5.16,

$$\mathrm{w}(\sigma) = x(x^{-1}x)^{k_0}u_1(x^{-1}x)^{k_1}u_2 \cdots u_h(x^{-1}x)^{k_h}x^{-1}$$

However, $u_i^2 = (\mathrm{w}(\sigma_i))^2 =_F \mathrm{w}(\sigma_i) = u_i$ by the Claim. That is, each u_i is idempotent. Hence, since $x^{-1}x$ is also an idempotent, and idempotents commute,

$$\mathrm{w}(\sigma) =_F x(x^{-1}x)^{k_0+k_1+\cdots+k_h}u_1u_2 \cdots u_hx^{-1} =_F x\overline{u}x^{-1},$$

where $\overline{u} = u_1u_2 \cdots u_h$. Furthermore, we have $\overline{u} = \mathrm{w}(\overline{\sigma})$, where $\overline{\sigma} = \sigma_1\sigma_2 \cdots \sigma_h$. Note that $\overline{\sigma}$ is a (ξ, ξ)-walk. Moreover, since σ spans T, it follows that $\overline{\sigma}$ spans $\overline{T}$.

Similarly, $\mathrm{w}(\tau) =_F x\overline{v}x^{-1}$, where $\overline{v} = \mathrm{w}(\overline{\tau})$ for some spanning (ξ, ξ)-walk $\overline{\tau}$ of $\overline{T}$. But $|V(\overline{T})| = n-1$ and so $\overline{u} =_F \overline{v}$ by the inductive hypothesis. Hence

$$\mathrm{w}(\sigma) =_F x\overline{u}x^{-1} =_F x\overline{v}x^{-1} =_F \mathrm{w}(\tau).$$

- $\gamma \neq \beta$. Then γ is a vertex of $\overline{T}$. Therefore, for some (ξ, ξ)-walks $\sigma_1, \sigma_2, \dots, \sigma_h$ and a (ξ, γ)-walk σ_∞ on $\overline{T}$ and some $k_i \in \mathbb{N} \cup \{0\}$ (where $i = 0, \dots, h$),

$$\sigma = \pi(\pi^{-1}\pi)^{k_0}\sigma_1(\pi^{-1}\pi)^{k_1}\sigma_2 \cdots \sigma_h(\pi^{-1}\pi)^{k_h}\sigma_\infty.$$

Let $u_i = \mathrm{w}(\sigma_i)$ for $i = 1, \dots, h$ and $u_\infty = \mathrm{w}(\sigma_\infty)$. By Lemma 5.16,

$$\mathrm{w}(\sigma) = x(x^{-1}x)^{k_0}u_1(x^{-1}x)^{k_1}u_2 \cdots u_h(x^{-1}x)^{k_h}u_\infty$$

By the Claim, $u_i^2 =_F u_i$ is idempotent for $i = 1, \dots, h$. Hence

$$\mathrm{w}(\sigma) =_F x(x^{-1}x)^{k_0+k_1+\cdots+k_h}u_1u_2 \cdots u_hu_\infty = x\overline{u},$$

where $\overline{u} = u_1u_2 \cdots u_hu_\infty$. Furthermore, we have $\overline{u} = \mathrm{w}(\overline{\sigma})$, where $\overline{\sigma} = \sigma_1\sigma_2 \cdots \sigma_h\sigma_\infty$ is a (ξ, γ)-walk that spans $\overline{T}$ since σ spans T. Similarly, $\mathrm{w}(\tau) =_F x\overline{v}$, where $\overline{v} = \mathrm{w}(\overline{\tau})$ for some spanning (ξ, ξ)-walk $\overline{\tau}$ of $\overline{T}$. By the inductive hypothesis, $\overline{u} =_F \overline{v}$ and so $\mathrm{w}(\sigma) = \mathrm{w}(\tau)$.

- β is not an endpoint of T. Then we can split T into two subtrees $\overline{T}_1$ and $\overline{T}_2$ such that $|V(\overline{T}_1)| < n$ and $|V(\overline{T}_2)| < n$, and $V(\overline{T}_1) \cap V(\overline{T}_2) = \{\beta\}$. Interchanging $\overline{T}_1$ and $\overline{T}_2$ if necessary, assume that $\gamma \in V(\overline{T}_2)$. Then

$$\sigma = \sigma_1\sigma_2 \cdots \sigma_h,$$

where h is even, the $\sigma_1, \sigma_3, \sigma_5, \dots, \sigma_{h-1}$ are (β, β)-walks (possibly trivial) on the subtree $\overline{T}_1$, the $\sigma_2, \sigma_4, \sigma_6, \dots, \sigma_{h-2}$ are (β, β)-walks (possibly trivial) on the subtree $\overline{T}_2$, and σ_h is a (β, γ)-walk (possibly trivial) on

the subtree $\overline{T}_2$. Let $u_i = \mathrm{w}(\sigma_i)$ for $i = 1, \dots, h$. Then for $i = 1, \dots, h-1$, by the Claim we have $u_i^2 =_F u_i$ and so u_i is an idempotent. Hence

$$\begin{aligned}\mathrm{w}(\sigma) &= u_1u_2u_3u_4 \cdots u_{h-1}u_h \\ &=_F u_1u_3 \cdots u_{h-1}u_2u_4 \cdots u_h \\ &= \overline{u}_1\overline{u}_2,\end{aligned}$$

where $\overline{u}_1 = \mathrm{w}(\overline{\sigma}_1)$ and $\overline{u}_2 = \mathrm{w}(\overline{\sigma}_2)$, and $\overline{\sigma}_k = \sigma_1\sigma_3 \cdots \sigma_{h-1}$ and $\overline{\sigma}_2 = \sigma_2\sigma_4 \cdots \sigma_h$. Note that $\overline{\sigma}_1$ is a (β, β)-walk on $\overline{T}_1$ and $\overline{\sigma}_2$ is a (β, γ)-walk on $\overline{T}_2$. Since σ spans T, it follows that $\overline{\sigma}_1$ spans $\overline{T}_1$ and $\overline{\sigma}_2$ spans $\overline{T}_2$.

Similarly, we can show that $\mathrm{w}(\tau) =_F \overline{v}_1\overline{v}_2$, where $\overline{v}_1 = \mathrm{w}(\overline{\tau}_1)$ and $\overline{v}_1 = \mathrm{w}(\overline{\tau}_1)$ for some spanning (β, β)-walk $\overline{\tau}_1$ of $\overline{T}_1$ and spanning (β, γ)-walk $\overline{\tau}_2$ of $\overline{T}_2$, respectively. Thus, by the inductive hypothesis, $\mathrm{w}(\overline{\sigma}_1) =_F \mathrm{w}(\overline{\tau}_1)$ and $\mathrm{w}(\overline{\sigma}_2) =_F \mathrm{w}(\overline{\tau}_2)$. Hence $\mathrm{w}(\sigma) = \overline{u}_1\overline{u}_2 =_F \overline{v}_1\overline{v}_2 =_F \mathrm{w}(\tau)$.

This completes the inductive step and so the result holds. Claim

Now we want to show each element of $\mathrm{FInvM}(A)$ is associated to a unique Munn tree. As a first step, the next lemma shows that each element of $(A \cup A^{-1})^*$ is associated to a unique Munn tree.

LEMMA 5.18. *Let T and $\overline{T}$ be Munn trees. Let σ be a spanning (α_T, ω_T)-walk in T and τ a spanning $(\alpha_{\overline{T}}, \omega_{\overline{T}})$-walk in $\overline{T}$ such that $\mathrm{w}(\sigma) = \mathrm{w}(\tau)$. Then T and $\overline{T}$ are isomorphic.*

Proof of 5.18. Let $x_1x_2 \cdots x_m = \mathrm{w}(\sigma) = \mathrm{w}(\tau)$ and suppose

$$\sigma = (\alpha_T = \delta_0, \dots, \delta_m = \omega_T) \quad \text{and} \quad \tau = (\alpha_{\overline{T}} = \eta_0, \dots, \eta_m = \omega_{\overline{T}}),$$

where x_i is the label on $\delta_{i-1}\delta_i$ and $\eta_{i-1}\eta_i$ for $i = 1, \dots, m$. Let T_i and $\overline{T}_i$ be the subtrees of T and $\overline{T}$ spanned by the walks $(\delta_0, \dots, \delta_i)$ and $(\eta_0, \dots, \eta_i)$, respectively, for $i = 0, \dots, m$. Notice that $T = T_m$ and $\overline{T} = \overline{T}_m$ since σ and τ span T and $\overline{T}$, respectively.

Clearly the map $\varphi_0 : T_0 \to \overline{T}_0$ defined by $\delta_0\varphi_0 = \eta_0$ is trivially an isomorphism of word trees.

Suppose that we have an isomorphism $\varphi_{i-1} : T_{i-1} \to \overline{T}_{i-1}$ such that $\delta_j\varphi_{i-1} = \eta_j$ for $j = 0, \dots, i-1$. We show that this can be extended to an isomorphism $\varphi_i : T_i \to \overline{T}_i$. We consider the cases $\delta_i \in V(T_{i-1})$ and $\delta_i \notin V(T_{i-1})$ separately.

- $\delta_i \in V(T_{i-1})$. Then $T_i = T_{i-1}$. Since δ_i is adjacent to δ_{i-1} and $\delta_{i-1}\delta_i$ has label x_i, there exists $\zeta \in V(\overline{T}_{i-1})$ such that $\zeta = \delta_i\varphi_{i-1}$ and ζ is adjacent to η_{i-1} with the edge $\eta_{i-1}\zeta$ having label x_i. However, η_i is adjacent to η_{i-1} in $\overline{T}$ and $\eta_{i-1}\eta_i$ has label x_i. Since $\overline{T}$ satisfies (5.13), $\eta_i = \zeta = \delta_i\varphi_{i-1}$. Thus $\overline{T}_i = \overline{T}_{i-1}$. So define $\varphi_i : T_i \to \overline{T}_i$ by $\varphi_i = \varphi_{i-1}$; then $\delta_j\varphi_i = \eta_j$ for $j = 0, \dots, i$ and so φ_i is an isomorphism of word trees.

- $\delta_i \notin V(T_{i-1})$. Suppose with the aim of obtaining a contradiction that $\eta_i \in \overline{T}_{i-1}$. Since η_i is adjacent to η_{i-1} and $\eta_{i-1}\eta_i$ has label x_i, there exists $\xi \in V(\overline{T}_{i-1})$ such that $\xi = \eta_i\varphi_{i-1}^{-1}$ and ξ is adjacent to δ_{i-1} with the edge $\delta_{i-1}\xi$ having label x_i. Since δ_i is adjacent to δ_{i-1} in T and $\delta_{i-1}\delta_i$ has label x_i, and T satisfies (5.13), we have $\delta_i = \xi \in (V(\overline{T}_{i-1}))\varphi_{i-1}^{-1} = V(T_{i-1})$, which is a contradiction. Hence $\eta_i \notin \overline{T}_{i-1}$.

  Thus we can extend φ_{i-1} to an isomorphism of word trees $\varphi_i : T_i \to \overline{T}_i$ by defining $\delta_i\varphi_i = \eta_i$.

By induction on i, there exists an isomorphism $\varphi_n : T_n \to \overline{T}_n$. Note that $\alpha_T\varphi_n = \delta_0\varphi_n = \eta_0\varphi_n = \alpha_{\overline{T}}\varphi_n$ and similarly $\omega_T\varphi_n = \omega_{\overline{T}}\varphi_n$. So φ_n is an isomorphism of Munn trees. 5.18

The next result strengthens the previous one, showing that each element of $\mathrm{FInvM}(A)$ is associated to a unique Munn tree.

LEMMA 5.19. *Let T and $\overline{T}$ be Munn trees. Let σ be a spanning (α_T, ω_T)-walk in T and τ a spanning $(\alpha_{\overline{T}}, \omega_{\overline{T}})$-walk in $\overline{T}$ such that $\mathrm{w}(\sigma) =_F \mathrm{w}(\tau)$. Then T and $\overline{T}$ are isomorphic.*

Proof of 5.19. It is sufficient to prove the result when $\mathrm{w}(\sigma)$ and $\mathrm{w}(\tau)$ differ by a single elementary ρ-transition.

- $\mathrm{w}(\sigma) = puq$ and $\mathrm{w}(\tau) = puu^{-1}uq$ for $p, u, q \in (A \cup A^{-1})^*$ with $u \neq \varepsilon$. So there exist (α_T, β)-, (β, γ)-, and (γ, ω_T)-walks σ_1, σ_2, and σ_3 on T such that $\sigma = \sigma_1\sigma_2\sigma_3$, where $\mathrm{w}(\sigma_1) = p$, $\mathrm{w}(\sigma_2) = u$, and $\mathrm{w}(\sigma_3) = q$. Let υ be the (α_T, ω_T)-walk $\sigma_1\sigma_2\sigma_2^{-1}\sigma_2\sigma_3$. Since σ spans T, so does υ. By Lemma 5.16, $\mathrm{w}(\upsilon) = puu^{-1}uq = \mathrm{w}(\tau)$. Therefore, by Lemma 5.18, the Munn trees T and $\overline{T}$ are isomorphic.
- $\mathrm{w}(\sigma) = puu^{-1}\upsilon\upsilon^{-1}q$ and $\mathrm{w}(\tau) = p\upsilon\upsilon^{-1}uu^{-1}$ for $p, u, \upsilon, q \in (A \cup A^{-1})^*$ with $u, \upsilon \neq \varepsilon$. So there exist (α_T, β)-, (β, γ)-, (β, δ)- and (β, ω_T)-walks $\sigma_1, \sigma_2, \sigma_3$, and σ_4 on T such that $\sigma = \sigma_1\sigma_2\sigma_2^{-1}\sigma_3\sigma_3^{-1}\sigma_4$, where $\mathrm{w}(\sigma_1) = p$, $\mathrm{w}(\sigma_2) = u$, $\mathrm{w}(\sigma_3) = \upsilon$ and $\mathrm{w}(\sigma_4) = q$. Let υ be the (α_T, ω_T)-walk $\sigma_1\sigma_3\sigma_3^{-1}\sigma_2\sigma_2^{-1}\sigma_4$. Since σ spans T, so does υ. By Lemma 5.16, $\mathrm{w}(\upsilon) = p\upsilon\upsilon^{-1}uu^{-1}q = \mathrm{w}(\tau)$. Therefore, by Lemma 5.18, the Munn trees T and $\overline{T}$ are isomorphic. 5.19

THEOREM 5.20. *Let T and $\overline{T}$ be Munn trees, and let σ be a spanning (α_T, ω_T)-walk on T, and let τ be a spanning $(\alpha_{\overline{T}}, \omega_{\overline{T}})$ walk on $\overline{T}$. Then $\mathrm{w}(\sigma) =_F \mathrm{w}(\tau)$ if and only if T and $\overline{T}$ are isomorphic.*

Equal in $\mathrm{FInvM}(A)$ $\Leftrightarrow$ isomorphic Munn trees

Proof of 5.20. If $w(\sigma) = \mathrm{w}(\tau)$, then T and $\overline{T}$ are isomorphic by Lemma 5.19.

On the other hand, suppose $\varphi : T \to \overline{T}$ is an isomorphism. Then φ maps σ to a spanning $(\alpha_{\overline{T}}, \omega_{\overline{T}})$-walk $\overline{\sigma}$ of $\overline{T}$. Note that $\mathrm{w}(\sigma) = \mathrm{w}(\overline{\sigma})$. Then by Lemma 5.17, $\mathrm{w}(\overline{\sigma}) = \mathrm{w}(\tau)$ and so $\mathrm{w}(\sigma) = \mathrm{w}(\tau)$. 5.20

We can use Munn trees to compute multiplications in $\mathrm{FInvM}(A)$. Suppose we have two Munn trees T_1 and T_2. Pick a spanning $(\alpha_{T_1}, \omega_{T_1})$-walk σ_1 on T_1 and a spanning $(\alpha_{T_2}, \omega_{T_2})$-walk σ_1 with elements. Merge the

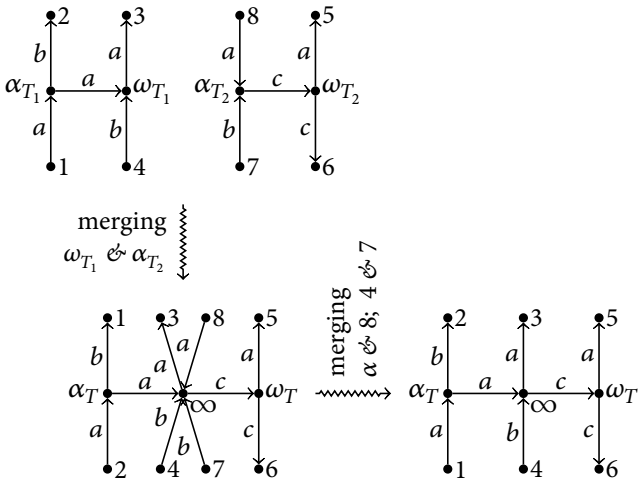


FIGURE 5.5
Multiplying using Munn trees: from Munn trees for $a^{-1}abb^{-1}a^2a^{-1}b^{-1}b$ and $caa^{-1}cc^{-1}b^{-1}ba^{-1}ac$, we compute a Munn tree for the product by merging the 'ω' of the first tree with the 'α' of the second and then folding edges.

the vertices ω_{T_1} and α_{T_2} to obtain a tree T, and let $\alpha_T = \alpha_{T_1}$ and $\omega_T = \omega_{T_2}$. Let $\sigma = \sigma_1\sigma_2$. Then σ is a spanning (α, ω)-walk on T. It remains to fold edges together until (5.13) is satisfied, as we did to construct Munn trees initially. Figure 5.5 illustrates the process.

EXERCISES

[See pages 225–234 for the solutions.]

*5.1 Let $X = \{1, 2\}$. Find a subsemigroup of $\mathcal{I}_X$ that contains only two elements and which is not an inverse subsemigroup.

5.2 Let G be a group and let S be the set of isomorphisms between subgroups of G. Prove that S is an inverse subsemigroup of $\mathcal{I}_G$.

5.3 Let X be a set and let $\sigma, \tau \in \mathcal{I}_X$. Prove the following:

a) $\sigma \mathcal{L} \tau \Leftrightarrow \operatorname{im}\sigma = \operatorname{im}\tau$;

b) $\sigma \mathcal{R} \tau \Leftrightarrow \operatorname{dom}\sigma = \operatorname{dom}\tau$;

c) $\sigma \mathcal{D} \tau \Leftrightarrow \sigma \mathcal{J} \tau \Leftrightarrow |\operatorname{dom}\sigma| = |\operatorname{dom}\tau|$.

*5.4 Let $X = \{1, \ldots, n\}$ with $n \geqslant 3$. Let $\tau = (1\ 2)$ and $\zeta = (1\ 2\ \ldots\ n-1\ n)$. As remarked in Exercise 1.11, from elementary group theory, we know that $\mathcal{S}_X = \langle \tau, \zeta \rangle$. For $k = 1, \ldots, n$, let

$$J_k = \{\sigma \in \mathcal{I}_X : |\operatorname{dom}\tau| = k\}.$$

(Notice that $J_n = \mathcal{S}_X$.) Fix an element β of J_{n-1}.

a) Let $\gamma \in J_{n-1}$ and let $\pi : \operatorname{dom}\gamma \to \operatorname{dom}\beta$ be a bijection. Prove that there exists $\rho \in \mathcal{S}_X$ such that $\pi\beta\rho = \gamma$. Deduce that $J_{n-1} \subseteq \langle \tau, \zeta, \beta \rangle$.

b) Prove that $J_k \subseteq J_{k+1}J_{n-1}$ for $k = 0, 1, \ldots, n-2$. Deduce that $\mathcal{I}_X = \langle \tau, \zeta, \beta \rangle$.

5.5 Let X be an infinite set.

a) Let $\tau \in \mathcal{I}_X$ be such that $\operatorname{dom}\tau = X$ and $\operatorname{im}\tau \subsetneq X$. (So τ is a bijection from X to a proper subset of itself.) Prove that $\langle \tau, \tau^{-1} \rangle$ is isomorphic to the bicyclic monoid.

b) Let I be an abstract index set with $|I| \geqslant 2$ and let $\{ \tau_i : i \in I \} \subseteq \mathcal{I}_X$ be a collection of partial bijections such that $\operatorname{dom}\tau_i = X$ and all the images $\operatorname{im}\tau_i$ are disjoint. (So the τ_i are bijections from X to disjoint subsets of X.) Prove that $\langle \{ \tau_i, \tau_i^{-1} : i \in I \} \rangle$ is an inverse monoid isomorphic to the monoid

$$\left.\begin{aligned} \operatorname{Mon}\langle z, b_i, c_i \text{ for } i \in I \mid\ & (b_i c_i, \varepsilon), (b_i c_j, z), \\ & (b_i z, z), (z b_i, z), \\ & (c_i z, z), (z c_i, z), (zz, z) \\ & \text{for } i, j \in I \text{ with } i \neq j \rangle. \end{aligned}\right\} \tag{5.14}$$

[These monoids are called the *polycyclic monoids*.] Polycyclic monoid

∗5.6 A semigroup is *orthodox* if it is regular and its set of idempotents form a subsemigroup. Orthodox

a) Prove that a Clifford semigroup is orthodox.

b) Prove that a semigroup is completely simple and orthodox if and only if it is isomorphic to the direct product of a rectangular band and a group.

∗5.7 Prove that a completely 0-simple semigroup is inverse if and only if it is isomorphic to $\mathcal{M}_0[G; I, I; P]$ where P is a diagonal $I \times I$ matrix.

∗5.8 Let S be a cancellative semigroup. An element τ of $\mathcal{I}_{S^1}$ is a *partial right translation* if $\operatorname{dom}\tau$ is a left ideal of S^1 and for any $x \in \operatorname{dom}\tau$ and $y \in S^1$, we have $(yx)\tau = y(x\tau)$. Partial right translation

a) Prove that if $\tau \in \mathcal{I}_{S^1}$ is a partial right translation, then $\operatorname{im}\tau$ is a left ideal of S^1.

b) Note that for each $x \in S$, the map $\rho_x : S^1 \to S^1$ (where $t\rho_x = tx$) is injective and so lies in $\mathcal{I}_{S^1}$. Let $\varphi : S \to \mathcal{I}_S$ be the homomorphism defined by $x \mapsto \rho_x$. Let T be the inverse subsemigroup of $\mathcal{I}_{S^1}$ generated by $\operatorname{im}\varphi$. Prove that the set of partial right translations in $\mathcal{I}_{S^1}$ is an inverse subsemigroup of of $\mathcal{I}_{S^1}$ and contains T.

∗5.9 Prove that the bicyclic monoid $B = \operatorname{Mon}\langle b, c \mid (bc, \varepsilon) \rangle$ is an inverse semigroup. [Hint: use the characterization of idempotents in Exercise 2.10(a).]

∗5.10 Let S be an inverse semigroup, and let $x \in S$ and $e \in E(S)$. Prove that $x \preccurlyeq e \Rightarrow x \in E(S)$.

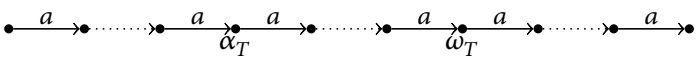
5.11 Prove that the $\operatorname{FInvM}(\{a\})$ is isomorphic to the set

$$K = \{ (p, q, r) : p, q, r \in \mathbb{Z}, p \leqslant 0, r \geqslant 0, p \leqslant q \leqslant r \}$$

with the operation

$$(p, q, r)(p', q', r') = (\min\{p, p' + q\}, q + q', \max\{r, q + r'\}).$$

[Hint: each element of FInvM($\{a\}$) corresponds to a Munn tree T of the form



View vertices along this path as having an 'x-coordinate' relative to α_T. Let p, q, and r be, respectively, the x-coordinates of the leftmost endpoint, the vertex ω_T, and the rightmost endpoint.]

5.12 Using Exercise 5.11 and the map

$$\varphi : K \to B \times B; \qquad (p, q, r)\varphi = (c^{-p}b^{-p+q}, c^r b^{-q+r}).$$

prove that FInvM($\{a\}$) is a subdirect product of two copies of the bicyclic monoid.

Bruck–Reilly extensions

5.13 Let M be a monoid presented by $\mathrm{Mon}\langle A \mid \rho\rangle$. Let $\varphi : M \to M$ be an endomorphism. The *Bruck–Reilly extension of M with respect to φ*, denoted $\mathrm{BR}(M, \varphi)$, is the monoid presented by

$$\left.\begin{array}{l}\mathrm{Mon}\langle A \cup \{b, c\} \mid \\ \qquad \rho \cup \{\, (bc, \varepsilon), (ba, (a\varphi)b), (ac, c(a\varphi)) : a \in A \,\}\rangle,\end{array}\right\} \quad (5.15)$$

where we view $a\varphi$ in the defining relations as some word in A^* representing that element of M.

a) Prove that every element of $\mathrm{BR}(M\varphi)$ is represented by a word of the form $c^\gamma w b^\beta$, where $\gamma, \beta \in \mathbb{N} \cup \{0\}$ and $w \in A^*$.

b) i) Prove that if $\gamma = \gamma'$, and $\beta = \beta'$, and $w =_M w'$, then we have $c^\gamma w b^\beta =_{\mathrm{BR}(M,\varphi)} c^{\gamma'} w' b^{\beta'}$.

ii) Let

$$\begin{aligned} X &= (\mathbb{N} \cup \{0\}) \times M \times (\mathbb{N} \cup \{0\}) \\ &= \{\, (\gamma, w, \beta) : \gamma, \beta \in \mathbb{N} \cup \{0\}, w \in M \,\}. \end{aligned}$$

Define

$$\begin{aligned} (\gamma, w, \beta)\tau_a &= (\gamma, w(a\varphi^\beta), \beta) \qquad \text{for each } a \in A; \\ (\gamma, w, \beta)\tau_b &= (\gamma, w, \beta + 1); \\ (\gamma, w, \beta)\tau_c &= \begin{cases} (\gamma + 1, w\varphi, 0) & \text{if } \beta = 0, \\ (\gamma, w, \beta - 1) & \text{if } \beta > 0. \end{cases} \end{aligned}$$

Prove that the map $\psi : A \to \mathcal{T}_X$ given by $x\psi = \tau_x$ for all $x \in A \cup \{b, c\}$ extends to a well-defined homomorphism ψ :

$\mathrm{BR}(M,\varphi) \to \mathcal{T}_X$. Prove that the homomorphism ψ is injective. Deduce that if $c^\gamma w b^\beta =_{\mathrm{BR}(M,\varphi)} c^{\gamma'} w' b^{\beta'}$, then $\gamma = \gamma'$, $\beta = \beta'$, and $w =_M w'$.

[Note that, since ψ is injective, $\mathrm{BR}(M,\varphi)$ is isomorphic to a particular subsemigroup of $\mathcal{T}_X$. Since this subsemigroup is independent of the choice of the presentation $\mathrm{Mon}\langle A \mid \rho\rangle$ for M, the Bruck–Reilly extension $\mathrm{BR}(M,\varphi)$ is also is independent of the choice of the presentation for M.]

c) Deduce that M embeds into $\mathrm{BR}(M,\varphi)$.

5.14 Let M be a monoid and let $\varphi : M \to M$ be defined by $x\varphi = 1$ for all $x \in M$. Prove that $\mathrm{BR}(M,\varphi)$ is simple. [Thus, as a consequence of Exercise 5.13, every semigroup S embeds into a simple semigroup $\mathrm{BR}(S^1,\varphi)$.]

NOTES

The exposition of the Vagner–Preston representation theorem is based on Clifford & Preston, *The Algebraic Theory of Semigroups*, § 1.9 and Howie, *Fundamentals of Semigroup Theory*, § 5.1. The discussion of Clifford semigroups is based on Howie, *Fundamentals of Semigroup Theory*, §§ 4.1–2. ◆ The introduction of free inverse semigroups follows Lawson, *Inverse Semigroups*, ch. 6; the explanation of Munn trees follows closely Munn, 'Free Inverse Semigroups' (which is a model of clarity) except that we consider free inverse monoids rather than free inverse semigroups. ◆ See Clifford & Preston, *The Algebraic Theory of Semigroups*, p. 28 for the quotation in the introduction. ◆ The Vagner–Preston theorem (Theorem 5.8), and much of the basic theory of inverse semigroups, is found in Vagner, 'Generalized groups' and Preston, 'Inverse semi-groups with minimal right ideals'; Preston, 'Representations of inverse semi-groups'. The structure theorem for Clifford semigroups is due to Clifford, 'Semigroups admitting relative inverses', though the terminology is later. ◆ For further reading, the standard text on inverse semigroups remains Petrich, *Inverse Semigroups*, but Lawson, *Inverse Semigroups* provides a geometric and topological perspective.

Commutative semigroups

6

‘The two operations, suicide and going to MIT, didn’t commute’

— Murray Gell-Mann,
‘The Making of a Physicist’. In: Edge.org.

Abelian groups (that is, commutative groups) have a simpler structure and are better understood than general groups, especially in the finitely generated case. It is therefore unsurprising that commutative semigroups also have a well-developed theory. However, there are still many more commutative semigroups than abelian groups. For instance, there are three essentially different (non-isomorphic) abelian groups with 8 elements (the cyclic group C_8 and the direct products $C_4 \times C_2$ and $C_2 \times C_2 \times C_2$), but there are 221 805 non-isomorphic commutative semigroups with 8 elements.

A large theory of commutative semigroups has developed, and we will sample only two areas: first, in structure theory, how cancellative commutative semigroups are group-embeddable; second, free commutative semigroups and their congruences, leading to the result that finitely generated semigroups are always finitely presented.

CANCELLATIVE COMMUTATIVE SEMIGROUPS

Example 2.14 showed that a cancellative semigroup is not necessarily group-embeddable. However, in this section we will see that a cancellative *commutative* semigroup is always group-embeddable. The method used to construct the group from the cancellative semigroup is essentially the same as that used to construct a field from an integral domain (for example, to construct $(\mathbb{Q}, +, \cdot)$ from $(\mathbb{Z}, +, \cdot)$).

Cancellative commutative semigroups are group-embeddable.

THEOREM 6.1. *Let S be a cancellative commutative semigroup. Then S embeds into a group G via a monomorphism $\varphi : S \to G$ such that $G = (S\varphi)(S\varphi)^{-1} = \{xy^{-1} : x, y \in S\}$.*

Proof of 6.1. First of all, note that S embeds into S^1 and that if $G = SS^{-1}$, then $G = S^1(S^1)^{-1}$, and so we assume without loss of generality that S is a monoid.

Define a relation σ on $S \times S$ by $(x, y)\ \sigma\ (z, t) \Leftrightarrow xt = zy$. It is trivial to prove σ is reflexive and symmetric since S is commutative, and σ is transitive since

$$\begin{aligned}
&(x, y)\ \sigma\ (z, t) \wedge (z, t)\ \sigma\ (p, q)\\
\Rightarrow\ &xt = zy \wedge zq = pt\\
\Rightarrow\ &xtzq = zypt\\
\Rightarrow\ &xq = py && \text{[since } S \text{ is cancellative and commutative]}\\
\Rightarrow\ &(x, y)\ \sigma\ (p, q).
\end{aligned}$$

Thus σ is an equivalence relation. Furthermore,

$$\begin{aligned}
&(x, y)\ \sigma\ (z, t) \wedge (x', y')\ \sigma\ (z', t')\\
\Rightarrow\ &xt = zy \wedge x't' = z'y'\\
\Rightarrow\ &xtx't' = zyz'y'\\
\Rightarrow\ &xx'tt' = zz'yy' && \text{[since } S \text{ is commutative]}\\
\Rightarrow\ &(xx', yy')\ \sigma\ (zz', tt').
\end{aligned}$$

Thus σ is a congruence.

Let $G = (S \times S)/\sigma$. Let $[(x, y)]_\sigma \in G$; then $(1_S x)y = (1_S y)x$ since S is commutative. Hence $(1_S x, 1_S y)\ \sigma\ (x, y)$ and thus $[(1_S, 1_S)]_\sigma[(x, y)]_\sigma = [(1_S x, 1_S y)]_\sigma = [(x, y)]_\sigma$. Similarly, $[(x, y)]_\sigma[(1_S, 1_S)]_\sigma = [(x, y)]_\sigma$. So G is a monoid with identity $[(1_S, 1_S)]_\sigma$.

Furthermore, $1_S(xy) = (yx)1_S$, since S is commutative, and therefore $(xy, yx)\ \sigma\ (1_S, 1_S)$. Hence $[(x, y)]_\sigma[(y, x)]_\sigma = [(xy, yx)]_\sigma = [(1_S, 1_S)]_\sigma$ and similarly $[(y, x)]_\sigma[(x, y)]_\sigma = [(1_S, 1_S)]_\sigma$. Thus $[(y, x)]_\sigma$ is a left and right inverse for $[(x, y)]_\sigma$. So G is a group.

Let $\varphi : S \to G$ be defined by $s\varphi = [(s, 1_S)]_\sigma$. It is clear that φ is a homomorphism. Furthermore, φ is injective since

$$\begin{aligned}
x\varphi = y\varphi &\Rightarrow [(x, 1_S)]_\sigma = [(y, 1_S)]_\sigma\\
&\Rightarrow s1_S = t1_S\\
&\Rightarrow s = t.
\end{aligned}$$

Therefore S embeds into G. Finally, note that

$$[(x, y)]_\sigma = [(x, 1_S)]_\sigma[(1_S, y)]_\sigma = (x\varphi)(y\varphi)^{-1} \in (S\varphi)(S\varphi)^{-1};$$

hence $G = (S\varphi)(S\varphi)^{-1}$. 6.1

Let S be a commutative cancellative semigroup. By Theorem 6.1, there is a monomorphism φ from S into a group G such that $G = (S\varphi)(S\varphi)^{-1}$. We can therefore identify S with a subsemigroup of G such that $G = SS^{-1}$. For any commutative cancellative semigroup S, denote by $G(S)$ some fixed group containing S as a subsemigroup, such that $G(S) = SS^{-1}$. (Actually, one can prove that $G(S)$ is unique up to isomorphism, but we will not need this result.)

FREE COMMUTATIVE SEMIGROUPS

Let A be an alphabet. Let $\mathrm{FCommS}(A)$ be semigroup presented by $\mathrm{Sg}\langle A \mid \rho \rangle$, where

$$\rho = \{ (ab, ba) : a, b \in A \}.$$

The following result is essentially immediate:

PROPOSITION 6.2. *$\mathrm{FCommS}(A)$ is a commutative semigroup.* 6.2

Free commutative semigroups

Let F be a commutative semigroup, let A be an alphabet, and let $\iota : A \to F$ be an embedding of A into F. Then the commutative semigroup F is the *free commutative semigroup on A* if, for any commutative semigroup S and map $\varphi : A \to S$, there is a unique homomorphism $\overline{\varphi} : F \to S$ that extends φ (that is, with $\iota\overline{\varphi} = \varphi$). Using diagrams, this definition says that F is a free commutative semigroup on A if

$$\text{for all } \begin{array}{ccc} A & \xhookrightarrow{\iota} & F \\ & {\scriptstyle\varphi}\searrow & \\ & & S \end{array} \text{ with } S \text{ commutative, there exists a unique homomorphism } \overline{\varphi} \text{ such that } \begin{array}{ccc} A & \xhookrightarrow{\iota} & F \\ & {\scriptstyle\varphi}\searrow & \downarrow{\scriptstyle\overline{\varphi}} \\ & & S \end{array}. \tag{6.1}$$

This definition is analogous to the definition of the free semigroup on A (see pages 38 f.) and free inverse semigroup on A (see page 107). As already noted, in Chapter 8, we will see definitions of 'free objects' in a much more general setting.

Reasoning analogous to the proof of Proposition 5.15 establishes the following result:

PROPOSITION 6.3. *Let A be an alphabet and let F be a commutative semigroup. Then F is a free commutative semigroup on A if and only if $F \simeq \mathrm{FCommS}(A)$.* 6.3

Uniqueness of the free commutative semigroup on A

Free commutative monoids

As in the discussions of 'free' and 'free inverse', we could repeat the reasoning above, but for monoids instead of semigroups. The monoid $\mathrm{FCommM}(A)$ is presented by $\mathrm{Mon}\langle A \cup A^{-1} \mid \rho \rangle$. A monoid F is a *free commutative monoid on A* if, for any commutative monoid S and map $\varphi : A \to S$, there is a unique monoid homomorphism $\overline{\varphi} : F \to S$ extending φ, with $\iota\overline{\varphi} = \varphi$. A commutative monoid F is a free commutative monoid on A if and only if $F \simeq \mathrm{FCommM}(A)$. We have $\mathrm{FCommM}(A) \simeq (\mathrm{FCommS}(A))^1$.

PROPOSITION 6.4. *Let A be a finite alphabet. Then* $\mathrm{FCommM}(A) \simeq (\mathbb{N} \cup \{0\})^A$.

(Recall the notation for cartesian and direct products from page 4.)

Proof of 6.4. Following Method 2.9, we aim to prove that $\mathrm{Mon}\langle A \mid \rho\rangle$ presents $(\mathbb{N} \cup \{0\})^A$. Define a map $\varphi : A \to (\mathbb{N} \cup \{0\})^A$, where $a\varphi$ is such that $(a)(a\varphi) = 1$ and $(x)(a\varphi) = 0$ for $x \neq a$. (That is, $a\varphi$ is the tuple whose a-th component is 1 and all other components 0.) Clearly $(\mathbb{N} \cup \{0\})^A$ satisfies the defining relations in ρ with respect to φ. Suppose $A = \{a_1, \ldots, a_n\}$ and let

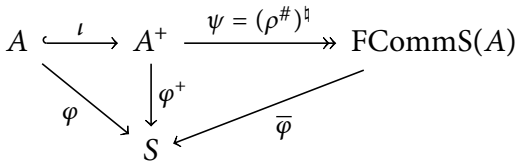
$$N = \{\, a_1^{k_1} a_2^{k_2} \cdots a_n^{k_n} : k_1, k_2, \ldots, k_n \in \mathbb{N} \cup \{0\} \,\}.$$

It is obvious that every word in A^* can be transformed to one in N by applying defining relations from ρ. Finally, since $(a_i)((a_1^{k_1} a_2^{k_2} \cdots a_n^{k_n})\varphi)$, we see that $\varphi|_N$ is injective. So $\mathrm{Mon}\langle A \mid \rho\rangle$ presents $(\mathbb{N} \cup \{0\})^A$. 6.4

Proposition 6.4 does not hold if A is infinite. The tuples $a\varphi$ as defined in the proof do not generate $(\mathbb{N} \cup \{0\})^A$ when A is infinite, because no (finite) product of these tuples is equal to (for example) the tuple with all components 1.

PROPOSITION 6.5. *Let S be a finite generated commutative semigroup (respectively, commutative monoid), let $\varphi : A \to S$ be an assignment of generators (with A finite), and let $\overline{\varphi} : \mathrm{FCommS}(A) \to S$ (respectively, $\overline{\varphi} : \mathrm{FCommM}(A) \to S$) be the homomorphism extending φ. Suppose there is a finite set $\sigma \subseteq \ker \overline{\varphi}$ such that $\sigma^{\#} = \ker \overline{\varphi}$. Then S is finitely presented.*

Proof of 6.5. We prove the result for semigroups; the reasoning for monoids is similar. Let $\varphi^+ : A^+ \to S$ be the homomorphism extending φ. For brevity, let ψ be the natural homomorphism $(\rho^{\#})^{\natural} : A^+ \to \mathrm{FCommS}(A)$, where $a\psi = a(\rho^{\#})^{\natural} = [a]_{\rho^{\#}}$. Note that $\ker \psi = \ker(\rho^{\#})^{\natural}\rho^{\#}$. The following diagram commutes:



To show that S is finitely presented, we must find a finite subset of $\mathcal{B}_{A^+}$ that generates $\ker \varphi^+$.

For each $(x, y) \in \sigma$, fix $w_x \in x((\rho^{\#})^{\natural})^{-1}$ and $w_y \in y((\rho^{\#})^{\natural})^{-1}$ and let $\widehat{\sigma} = \{\, (w_x, w_y) : (x, y) \in \sigma \,\}$. Note that $\widehat{\sigma}$ is finite since σ is finite. We will prove that $(\widehat{\sigma} \cup \rho)^{\#} = \ker(\psi\overline{\varphi}) = \ker \varphi^+$.

Make the following definition: for any $\tau \in \mathcal{B}_T$, let $\tau\psi^{-1} = \{\, (s, t) \in S \times S : (s\psi, t\psi) \in \tau \,\}$.

Let $(u, v) \in (\sigma^{\mathsf{C}})\psi^{-1}$. So $(u\psi, v\psi) \in \sigma^{\mathsf{C}}$. So by Proposition 1.27, there exist $p, q \in \mathrm{FCommS}(A)$ and $(x, y) \in \sigma$ such that $u\psi = pxq$ and

$v\psi = pyq$. Let $p', q' \in S$ be such that $p'\psi = p$ and $q'\psi = q$. Then $(p'w_x q')\psi = u\psi$ and $(p'w_y q')\psi = v\psi$. Therefore $(u, p'w_x q') \in \ker\psi = \rho^\#$ and $(p'w_y q, v) \in \ker\psi = \rho^\#$. Since $(p'w_x q', p'w_y q') \in \widehat{\sigma}^\#$, we have

$$u \;\rho^\#\; p'w_x q' \;\widehat{\sigma}^\#\; p'w_y q' \;\rho^\#\; v,$$

and so $(u, v) \in (\widehat{\sigma} \cup \rho)^\#$. Thus $(\sigma^{\mathsf{C}})\psi^{-1} \subseteq (\widehat{\sigma} \cup \rho)^\#$. Since $(\widehat{\sigma} \cup \rho)^\#$ is symmetric, $(\sigma^{\mathsf{C}})\psi^{-1} \cup (\sigma^{\mathsf{C}})^{-1}\psi^{-1} \subseteq (\widehat{\sigma} \cup \rho)^\#$.

Now let $u, v \in A^+$. Then

$$
\begin{aligned}
&(u, v) \in \ker(\psi\overline{\varphi}) \\
\Rightarrow\; &(u\psi, v\psi) \in \sigma^\# \\
\Rightarrow\; &(u\psi = v\psi) \vee (u\psi, v\psi) \in \textstyle\bigcup_{n=1}^{\infty} (\sigma^{\mathsf{C}} \cup (\sigma^{\mathsf{C}})^{-1})^n \\
&\hfill \text{[by Proposition 1.26(f)]} \\
\Rightarrow\; &(\exists n \in \mathbb{N} \cup \{0\})(\exists w_0, \ldots, w_n \in \mathrm{FCommS}(A)) \\
&\qquad \big[(u\psi = w_0) \wedge (w_n = v\psi) \\
&\qquad\qquad \wedge (\forall i)\big((w_i, w_{i+1}) \in \sigma^{\mathsf{C}} \cup (\sigma^{\mathsf{C}})^{-1}\big)\big] \\
\Rightarrow\; &(\exists n \in \mathbb{N} \cup \{0\})(\exists w'_0, \ldots, w'_n \in A^+) \\
&\qquad \big[(u\psi = w'_0\psi) \wedge (w'_n\psi = v\psi) \\
&\qquad\qquad \wedge (\forall i)\big((w'_i, w'_{i+1}) \in \sigma^{\mathsf{C}}\psi^{-1} \cup (\sigma^{\mathsf{C}})^{-1}\psi^{-1}\big)\big] \\
&\hfill \text{[since } \psi \text{ is surjective]} \\
\Rightarrow\; &(\exists n \in \mathbb{N} \cup \{0\})(\exists w'_0, \ldots, w'_n \in A^+) \\
&\qquad \big[(u\psi = w'_0\psi) \wedge (w'_n\psi = v\psi) \\
&\qquad\qquad \wedge (\forall i)\big((w'_i\psi, w'_{i+1}) \in (\sigma' \cup \rho)^\#\big)\big] \\
&\hfill \text{[since } (\sigma^{\mathsf{C}})\psi^{-1} \cup (\sigma^{\mathsf{C}})^{-1}\psi^{-1} \subseteq (\widehat{\sigma} \cup \rho)^\#\text{]} \\
\Rightarrow\; &(u, v) \in (\sigma' \cup \rho)^\#)\big].
\end{aligned}
$$

Therefore $\ker(\psi\overline{\varphi}) \subseteq (\sigma' \cup \rho)^\#$.

On the other hand, if $(u, v) \in \sigma'$, then $(u\psi, v\psi) \in \sigma \subseteq \ker\overline{\varphi}$, so $u\psi\overline{\varphi} = v\psi\overline{\varphi}$ and so $(u, v) \in \ker(\psi\overline{\varphi})$. If $(u, v) \in \rho \subseteq \ker\psi$, then $u\psi = v\psi$, so $u\psi\overline{\varphi} = v\psi\overline{\varphi}$ and so $(u, v) \in \ker(\psi\overline{\varphi})$. Thus $\sigma' \cup \rho \subseteq \ker(\psi\overline{\varphi})$ and hence $(\sigma' \cup \rho)^\# \subseteq \ker(\psi\overline{\varphi})$ since $\ker(\psi\overline{\varphi})$ is a congruence.

Therefore $(\sigma' \cup \rho)^\# = \ker(\psi\overline{\varphi}) = \ker\varphi^+$ and so S is defined by the finite presentation $\mathrm{Sg}\langle A \mid \sigma' \cup \rho\rangle$. 6.5

Rédei's theorem

When the alphabet A has n elements, we write write $F_n = \mathrm{FCommM}(A)$ for brevity and (by Proposition 6.4) we view elements of

FCommM(A) as n-tuples of non-negative integers from $\mathbb{N} \cup \{0\}$. Define a relation $\leqslant$ on F_n by

$$(x_1, \dots, x_n) \leqslant (y_1, \dots, y_n) \Leftrightarrow (\forall i \in \{1, \dots, n\})(x_i \leqslant y_i).$$

It is easy to see that $\leqslant$ is a partial order on F_n. Notice that there are no infinite $\leqslant$-decreasing sequences in F_n.

Dickson's theorem

THEOREM 6.6. *Every antichain in F_n is finite.*

Proof of 6.6. The proof is by induction on n. For the base of the induction, let Y be an antichain in $\mathbb{N} \cup \{0\} \simeq F_1$. Since $\leqslant$ is a total order on $\mathbb{N} \cup \{0\}$, every pair of elements is comparable and thus Y contains at most one element. Thus the result holds for $n = 1$.

Now suppose the result holds for all $n < k$; we aim to prove it for $n = k$. Let $Y \subseteq (\mathbb{N} \cup \{0\})^k \simeq F_k$ be an antichain. For each $t \in \mathbb{N} \cup \{0\}$ and for each $i \in \{1, \dots, k\}$, let

$$Y_{i,t} = \{ (x_1, \dots, x_k) \in Y : x_i = t \}.$$

The set

$$\{ (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_k) : (x_1, \dots, x_{i-1}, t, x_{i+1}, x_k) \in Y_{i,t} \}.$$

is an antichain (possibly empty) of F_{k-1} and therefore finite by the induction hypothesis. Hence each set $Y_{i,t}$ is finite.

Fix some $y \in Y$, with $y = (y_1, \dots, y_k)$. Let $z = (z_1, \dots, z_k) \in Y \setminus \{y\}$. Since Y is an antichain, $y \not\leqslant z$. Hence $y_i > z_i$ for some $j \in \{i, \dots, k\}$. Hence

$$\begin{aligned} Y &= \{y\} \cup \bigcup_{i=1}^{n} \{ (z_1, \dots, z_k) \in Y : z_i < y_i \} \\ &= \{y\} \cup \bigcup_{i=1}^{n} \bigcup_{t=0}^{y_j - 1} \{ (z_1, \dots, z_k) \in Y : z_i = t \} \\ &= \{y\} \cup \bigcup_{i=1}^{n} \bigcup_{t=0}^{y_j - 1} Y_{i,t}. \end{aligned}$$

Each set $Y_{i,t}$ is finite, and so Y itself is finite. Since Y was an arbitrary antichain in F_k, this establishes the induction step and so proves the result. 6.6

PROPOSITION 6.7. *Let σ be a congruence on F_n. Then there is a finite set $\rho \subseteq \sigma$ such that $\rho^{\#} = \sigma$.*

Proof of 6.7. Define the lexicographic order $\sqsubseteq$ on F_n by

$$(x_1, \dots, x_n) \sqsubset (y_1, \dots, y_n) \Leftrightarrow (\exists k \in \{1, \dots, n\}) \\ \big(x_k < y_k \wedge (\forall j < k)(x_j = y_k)\big).$$

Then $\sqsubseteq$ is a total order on F_n and is compatible. (That is, $x \sqsubseteq y \Rightarrow xz \sqsubseteq yz$ for all $x, y, z \in F_n$.) Furthermore, $\sqsubseteq$ is a well-order (that is, every non-empty subset of F_n has a $\sqsubseteq$-minimum element). In particular, every σ-class $[x]_\sigma$ has a $\sqsubseteq$-minimum element q_x. Let

$$Q = \{ q_x : x \in F_n \} = \{ y \in F_n : (\forall x \in F_n)(y \,\sigma\, x \Rightarrow y \sqsubseteq x) \}.$$

So Q consists of the $\sqsubseteq$-minimal elements of all the σ-classes. Let R be the complement of Q in F_n; that is,

$$R = \{ x : q_x \sqsubset x \} = \{ x \in F_n : (\exists y \in F_n)(y \,\sigma\, x \wedge y \sqsubset x) \}.$$

Then R consists of the non-$\sqsubseteq$-minimal elements of all the σ-classes. Furthermore,

$$(x \in R) \wedge (z \in F_n) \Rightarrow (q_x \sqsubset x) \wedge (z \in S) \Rightarrow q_x z \sqsubset xz \Rightarrow xz \in R;$$

hence R is an ideal of F_n. Let M be the set of $\leqslant$-minimal elements of R. Then M is an antichain and so finite by Theorem 6.6. Let

$$\rho = \{ (q_m, m) : m \in M \}.$$

Notice that ρ is finite because M is finite.

The aim is now to show that $\rho^\# = \sigma$. Since $q_m \,\sigma\, m$ for each $m \in M$, it is immediate that $\rho \subseteq \sigma$ and so $\rho^\# \subseteq \sigma$.

To prove that $\sigma \subseteq \rho^\#$, the first step is to prove that $q_x \;\rho^\#\; x$ for all $x \in F_n$.

Suppose, with the aim of obtaining a contradiction, that $(q_x, x) \notin \rho^\#$ for some $x \in F_n$. Then, since $\sqsubseteq$ is a well-order, there is a $\sqsubseteq$-minimum $s \in F_n$ such that $(q_s, s) \notin \rho^\#$. This element s cannot be in Q, since otherwise $q_s = s$ and so $(q_s, s) \in \rho^\#$. Furthermore, s cannot be in M, since otherwise $(q_s, s) \in \rho$ by definition and hence $(q_s, s) \in \rho^\#$. Thus $s \in R \setminus M$ and so $s > m$ for some $m \in M$. Therefore $s = mt$ for some $t \in F_n$.

Let $u = q_m t$. Since $(q_m, m) \in \rho^\#$ and $(q_m, m) \in \sigma$, we have $(u, s) = (q_m t, mt) \in \rho^\#$ and so $(u, s) \in \sigma$. Notice that $(u, s) \in \sigma$ implies $q_u = q_s$. Furthermore, $u \sqsubset s$ since $q_m \sqsubset m$ and $\sqsubseteq$ is compatible. Since s is $\sqsubseteq$-minimum with $(q_s, s) \notin \rho^\#$, it follows that $(q_u, u) \in \rho^\#$. Therefore $s \;\rho^\#\; u \;\rho^\#\; q_x = q_s$. Thus $(q_s, s) \in \rho^\#$, which is a contradiction, and hence $(q_x, x) \in \rho^\#$ for all $x \in F_n$.

Finally, let $(x, y) \in \sigma$. Then $q_x = q_y$. By the previous paragraph, (q_x, x) and (q_y, y) are in $\rho^\#$. Thus $x \;\rho^\#\; q_x = q_y \;\rho^\#\; y$; hence $(x, y) \in \rho^\#$. That is, $\sigma \subseteq \rho^\#$, and therefore $\sigma = \rho^\#$. 6.7

The following result is immediate from Propositions 6.5 and Proposition 6.7:

RÉDEI'S THEOREM 6.8. *Every finitely generated commutative monoid is finitely presented.* 6.8

Rédei's theorem

EXERCISES

[See pages 234–237 for the solutions.]

∗6.1 Let S be a commutative semigroup. Let G and G' be abelian groups such that $G = SS^{-1}$ and $G' = SS^{-1}$. Prove that there is an isomorphism $\psi : G \to G'$ such that $\psi|_S$ maps $S \subseteq G$ to $S \subseteq G'$.

∗6.2 Let S be a commutative semigroup. Let I be an ideal of S, and let G be an abelian group. Let $\varphi : I \to G$ be a homomorphism. Prove that there is a unique extension of φ to a homomorphism $\widehat{\varphi} : S \to G$.

6.3 Let S be a non-trivial subsemigroup of $(\mathbb{N} \cup \{0\}, +)$. Prove that there exists $d \in \mathbb{N} \cup \{0\}$ such that $S \subseteq d\mathbb{N}$ and $d\mathbb{N} \setminus S$ is finite.

6.4 Let S be a subsemigroup of $(\mathbb{Z}, +)$. Prove that either every element of S is non-negative, or every element of S is non-positive, or S is a subgroup.

Right- and left-reversibility

∗6.5 A semigroup S is *right-reversible* (respectively, *left-reversible*) if every two elements of S have a common left (respectively, right) multiple; that is, if for all $x, y \in S$, there exist $z, t \in S^1$ such that $zx = ty$ (respectively, $xz = yt$).

Ore's theorem

Let S be a cancellative right-reversible semigroup; the aim of this exercise is to prove that S is group-embeddable; this is Ore's theorem, and generalizes Theorem 6.1. Let $\varphi : S \to \mathcal{I}_S$ be the homomorphism defined by $x \mapsto \rho_x$. Let T be the inverse subsemigroup of $\mathcal{I}_S$ generated by $\operatorname{im}\varphi$. By Exercise 5.8(b), every element of T is a partial right translation. Define a relation $\sim$ on T by

$$\alpha \sim \beta \Leftrightarrow (\exists \delta \in T)\big((\delta \subseteq \alpha) \wedge (\delta \subseteq \beta)\big)$$

for all $\alpha, \beta \in T$. Notice that

$$\delta \subseteq \alpha \Leftrightarrow \big((\operatorname{dom}\delta \subseteq \operatorname{dom}\alpha) \wedge (\forall x \in \operatorname{dom}\delta)(x\delta = x\alpha)\big).$$

a) Prove that $\sim$ is an congruence.

b) Let $G = T/\!\sim$. Prove that G is a group.

c) Let $\alpha, \beta \in T$. Prove that $\alpha \circ \beta$ is not the empty relation, and so deduce that T does not contain the empty relation.

d) Let $\psi = \varphi \circ {\sim}^\natural$ (that is, $x\psi = [x\varphi]_\sim$). Prove that ψ is a monomorphism and so deduce that S is group-embeddable.

∗6.6 Let $S = (\mathbb{N} \cup \{0\}) \times (\mathbb{N} \cup \{0\})$ and define a multiplication on S by

$$(m, n)(p, q) = (m + p, 2^p n + q)$$

Check that this multiplication is associative, so that S is a semigroup. Prove that S is left reversible but not right reversible.

NOTES

The number of commutative semigroups of with 8 elements is from Grillet, *Commutative Semigroups*, p. 1. ◆ Rédei's theorem (Theorem 6.8) was first proved by Rédei, *The Theory of Finitely Generated Commutative Semigroups*; see also Clifford & Preston, *The Algebraic Theory of Semigroups*, § 9.3. The proof given here is from Grillet, 'A short proof of Rédei's theorem'. ◆ Ore's theorem (Exercise 6.5) is contained in a theorem about rings proved, using different terminology, in Ore, 'Linear equations in non-commutative fields'; the proof here is due to Rees, 'On the group of a set of partial transformations'. ◆ For further reading, Grillet, *Commutative Semigroups* is a comprehensive monograph, but with a very terse style, and Rosales & García-Sánchez, *Finitely Generated Commutative Monoids* is an accessible introduction to structural and computational aspects.

Finite semigroups

7

‘The known is finite, the unknown infinite; intellectually we stand on an islet in the midst of an illimitable ocean of inexplicability.’

— T.H. Huxley,
On the Reception of the Origin of Species.

In this chapter, we begin the detailed study of finite semigroups. Although Green's relations will play a role, other techniques are used to understand finite semigroups. In particular we will introduce the notion of divisibility, where one semigroup is a homomorphic image of a subsemigroup of another. The goal of this chapter is to prove the Krohn–Rhodes theorem, which says that every finite semigroup divides a wreath product of finite groups and finite aperiodic semigroups, which, as we shall see, are finite semigroups where all subgroups are trivial. This leads naturally into the classification of finite semigroups by means of pseudovarieties, which is the topic of next chapter.

GREEN'S RELATIONS AND IDEALS

As a consequence of Proposition 3.3, we know that the Green's relations $\mathcal{D}$ and $\mathcal{J}$ coincide for finite semigroups.

PROPOSITION 7.1. *Let M be a finite monoid with identity 1. Then $H_1 = L_1 = R_1 = D_1 = J_1$. Furthermore, H_1 is the group of units of M, and $M \setminus H_1$ is either empty or an ideal of M.*

Proof of 7.1. Let $x \in R_1$. Then there exists $y \in M^1 = M$ such that $xy = 1$. Since M is finite, $x^{m+k} = x^m$ for some $m, k \in \mathbb{N}$. Then $x^k = x^{m+k}y^m = x^m y^m = 1$, and so $yx = 1yx = x^k yx = x^{k-1}x = x^k = 1$ Hence $x \mathcal{H} 1$. Therefore $R_1 \subseteq H_1$. The opposite inclusion is obvious, so $R_1 = H_1$. Similarly $L_1 = H_1$. So D_1 contains only one $\mathcal{L}$-class and only one $\mathcal{R}$-class and so $D_1 = H_1$. Finally, $J_1 = H_1$ since $\mathcal{D} = \mathcal{J}$.

This reasoning also shows that H_1 is contained in the group of units of M. On the other hand, all elements of group of units of M are $\mathcal{H}$-related to 1, so the group of units of S is H_1.

For any $y \in M \setminus H_1 = M \setminus J_1$, we have $J_y < J_1$ by (3.2). So $M \setminus H_1 = I(1) = \{ y \in S : J_y < J_1 \}$, which is either empty or an ideal by Lemma 3.9. 7.1

PROPOSITION 7.2. *Let S and S' be finite semigroups and let $\varphi : S \to S'$ be a surjective homomorphism. Let G' be a maximal subgroup of S'. Then there is a maximal subgroup G of S such that $G\varphi = G'$.*

Proof of 7.2. Let G' be a maximal subgroup of S'. Then $T = G'\varphi^{-1}$ is a subsemigroup of S and $T\varphi = G'$. Since T is finite, it has a kernel; let $K = K(S)$, which is a simple ideal of T by Proposition 3.10. Since φ is surjective, $K\varphi$ is an ideal of the group G' and so $K\varphi = G'$. Since K is finite it is also completely simple by Proposition 4.10. So, by Theorem 4.11, $K \simeq \mathcal{M}[G; I, \Lambda; P]$ for some group G, index sets I and Λ, and matrix P over G. View $\varphi|_K$ as a surjective homomorphism from $\mathcal{M}[G; I, \Lambda; P]$ to G'. For each $i \in I$ and $\lambda \in \Lambda$, let $G_{i\lambda}$ be the subset $\{i\} \times G \times \{\lambda\}$ of $\mathcal{M}[G; I, \Lambda; P]$. Then $\mathcal{M}[G; I, \Lambda; P]$ is the union of the various $G_{i\lambda}$, and $G_{i\lambda}G_{j\mu} \subseteq G_{i\mu}$. In particular, every $G_{i\lambda}$ is a subgroup of T.

Let $G'_{i\lambda} = G_{i\lambda}\varphi$. Then each $G'_{i\lambda}$ is a subgroup of G', and $G'_{i\lambda}G'_{j\mu} \subseteq G'_{i\mu}$. In particular, $G'_{i\lambda}G'_{j\lambda} \subseteq G'_{i\lambda}$, which implies $G'_{j\lambda} = 1_{G'}G'_{j\lambda} \subseteq G'_{i\lambda}$. Similarly $G'_{i\lambda} \subseteq G'_{i\mu}$. Thus all the $G'_{i\lambda}$ are equal. Since φ is surjective, G' is the union of the $G'_{i\lambda}$ and thus equal to any one of the $G'_{i\lambda}$. Hence $G' = G_{i\lambda}\varphi$ for any $i \in I$ and $\lambda \in \Lambda$. 7.2

PROPOSITION 7.3. *Let S be a finite semigroup and let $x, y \in S$. If $x \mathcal{H} y$, then $y \in xG$ for some subgroup G of S.*

Proof of 7.3. Let H be an $\mathcal{H}$-class of S. Apply Proposition 7.2 to the natural surjective homomorphism $\sigma_H^\natural : \mathrm{Stab}(H) \to \Gamma(H)$ to see that $H = G\sigma_H^\natural$ for some subgroup G of $\mathrm{Stab}(H)$. By Proposition 3.24, we see that $y \in x \cdot \Gamma(H) = xG$ for all $x, y \in H$. 7.3

Aperiodic semigroups

A semigroup S is *aperiodic* if for every $x \in S$, there exists $n \in \mathbb{N}$ such that $x^n = x^{n+1}$.

Notice that aperiodic semigroup are actually *periodic*. For example, any semigroup of idempotents (such as a semilattice) satisfies $x = x^2$ and so is aperiodic.

Characterization of aperiodic finite semigroups

PROPOSITION 7.4. *Let S be a finite semigroup. The following are equivalent:*

a) *S is aperiodic.*

b) *all subgroups of S are trivial;*

c) *$\mathcal{H} = \mathrm{id}_S$;*

Proof of 7.4. Part 1 [a) $\Rightarrow$ b)]. Suppose S is aperiodic. Let G be a subgroup of S and let $x \in G$. Then $x^m = x^{m+1}$ for some $m \in \mathbb{N}$. Hence $x = 1_G$ by cancellativity in G. So G is trivial. Thus all subgroups of S are trivial.

Part 2 [b) $\Rightarrow$ c)]. Suppose that all subgroups of S are trivial. Let H be an $\mathcal{H}$-class of S. Then $\Gamma(H)$, which is a homomorphic image of a subgroup

of $\mathrm{Stab}(H)$, is trivial. Since $|H| = |\Gamma(H)|$, it follows that H is trivial by Proposition 7.3. Hence $\mathcal{H} = \mathrm{id}_S$.

Part 3 [c) $\Rightarrow$ a)]. Suppose that $\mathcal{H} = \mathrm{id}_S$. Let $x \in S$. Since S is finite, $x^m = x^{m+k}$ for some $m, k \in \mathbb{N}$. The set of elements $\{x^m, x^{m+1}, \ldots, x^{m+k-1}\}$ is a subgroup and so all its elements are $\mathcal{H}$-related. Since $\mathcal{H} = \mathrm{id}_S$, this set thus contains only one element, which implies $k = 1$. Hence $x^m = x^{m+1}$. Thus S is aperiodic. 7.4

We end this section by proving the following two results, although we will not use them until Chapter 9:

LEMMA 7.5. *Let S be a finite semigroup and let $n \geqslant |S|$. Then $S^n = SE(S)S$.*

Proof of 7.5. Let $e \in E(S)$. Then $SeS = Se^{n-2}S \subseteq S^n$. Thus $SE(S)S \subseteq S^n$.

Let $x \in S^n$. Then $x = x_1 \cdots x_n$, where $x_i \in S$. Suppose first that all the products $x_1 \cdots x_k$ for $k \leqslant n$ are distinct. Then every element of S is equal to some product $x_1 \cdots x_k$. Hence, since S, being finite, contains at least one idempotent, some product $e = x_1 \cdots x_k$ is idempotent. Hence $x = ex_{k+1} \cdots x_n = e^3 x_{k+1} \cdots x_n \in SeS \subseteq SE(S)S$. Now suppose that $x_1 \cdots x_k = x_1 \cdots x_\ell$ for some $k < \ell$. Then $x_1 \cdots x_k = x_1 \cdots x_k (x_{k+1} \cdots x_\ell)^i$ for all $i \in \mathbb{N}$. Let i be such that $e = (x_{k+1} \cdots x_\ell)^i$ is idempotent. Then $x_1 \cdots x_n = x_1 \cdots x_k e x_{\ell+1} \cdots x_n = x_1 \cdots x_k e^3 \in SeS \subseteq SE(S)S$. Thus $S^n \subseteq SE(S)S$. 7.5

LEMMA 7.6. *Let S be a finite semigroup and let $x, y \in S$.*

a) *If $x \mathcal{D} xy$, then $x \mathcal{R} xy$.*

b) *If $x \mathcal{D} yx$, then $x \mathcal{L} yx$.*

Proof of 7.6. We prove only part a); dual reasoning gives part b).

Suppose $x \mathcal{D} xy$. Since $\mathcal{D} = \mathcal{J}$ by Proposition 3.3, there exist $p, q \in S^1$ such that $pxyq = x$. Thus $p^n x (yq)^n$ for all $n \in \mathbb{N}$. Since S is finite, there exists $k \in \mathbb{N}$ such that $e = p^k$ is idempotent. Thus, in particular, $ex(yq)^k = x$, and also $ex = eex(yq)^k = ex(yq)^k = x$. Combining these gives $x(yq)^k = x$. So $x \mathcal{R} xy$. 7.6

SEMIDIRECT AND WREATH PRODUCTS

Semidirect product

Let S and T be semigroups and let T act on S from the left by endomorphisms; let $\varphi : T \to \mathrm{End}(S)$ be the anti-homomorphism corresponding to this left action. To avoid having to write extra brackets, we will write ${}^t s$ instead of $t \cdot s$. The *semidirect product of S and T with*

respect to φ is denoted $S \rtimes_\varphi T$ and is the cartesian product $S \times T$ with multiplication defined by

$$(s_1, t_1)(s_2, t_2) = (s_1 {}^{t_1}s_2, t_1 t_2). \tag{7.1}$$

This multiplication is associative (see Exercise 7.6) and so $S \rtimes_\varphi T$ is a semigroup. Notice that $S \rtimes_\varphi T$ has cardinality $|S||T|$.

Notice that for any semigroups S and T, we can take the trivial left action, where $t \cdot s = {}^t s = s$ for $t \in T$ and $s \in S$; this corresponds to the trivial anti-homomorphism $\varphi : T \to \mathrm{End}(S)$, with $y\varphi = \mathrm{id}_S$ for all $y \in T$. In this case, $(s_1, t_1)(s_2, t_2) = (s_1 s_2, t_1 t_2)$. Thus the direct product is a special case of the semidirect product.

Wreath product

Recall from page 4 that S^T is the direct product of copies of the set S indexed by T, or formally the set of maps from T to S. Define a left action of T on S^T by letting $y \cdot f = {}^y f$ be such that $(x)\,{}^y f = (xy)f$. This satisfies the definition of a left action: for all $y, z \in T$, we have $z \cdot (y \cdot f) = zy \cdot f$ since, for all $x \in T$,

$$(x)\big(z \cdot (y \cdot f)\big) = (x)\,{}^z({}^y f) = (xz)\,{}^y f = (xzy)f = (x)\,{}^{zy} f = zy \cdot f.$$

Let φ be the anti-homomorphism that corresponds to this action. The *wreath product of* S *and* T, denoted $S \wr T$, is the semidirect product $S^T \rtimes_\varphi T$. Thus the product in $S \wr T$ is

$$(f_1, t_1)(f_2, t_2) = (f_1 {}^{t_1}f_2, t_1 t_2).$$

Since this multiplication is derived from the multiplication in direct and semidirect products, we know it is associative. Hence $S \wr T$ is a semigroup. Notice that $S \wr T$ has cardinality $|S|^{|T|}|T|$.

Let S, T, U be finite semigroups. Then

$$\big|(S \wr T) \wr U\big| = |S \wr T|^{|U|}|U| = \big(|S|^{|T|}|T|\big)^{|U|}|U| = |S|^{|T||U|}|T|^{|U|}|U|$$

and

$$\big|S \wr (T \wr U)\big| = |S|^{|T \wr U|}|T \wr U| = |S|^{|T|^{|U|}|U|}|T|^{|U|}|U|.$$

Therefore the wreath product, as an operation on semigroups, is not associative.

PROPOSITION 7.7. *If M and N are monoids, then $M \wr N$ is a monoid with identity $(e, 1_N)$, where $e : N \to M$ is the constant map with $(x)e = 1_M$ for all $x \in N$.*

Proof of 7.7. Suppose M and N are monoids. Let . Then for any $(f, n) \in M \wr N$, we have

$$\begin{aligned} &(e, 1_N)(f, n) \\ &= (e\,{}^{1_N}f, 1_N n) \\ &= (f, n) \qquad [\text{since } (x)e\,{}^{1_N}f = (x)e(x1_N)f = 1_M(x)f = (x)f] \end{aligned}$$

and

$$\begin{aligned}
&(f,n)(e,1_N)\\
&= (f\,{}^{n}e, n1_N)\\
&= (f,n); \qquad [\text{since } (x)f\,{}^{n}e = (x)f(xn)\,{}^{n}e = (x)f1_M = (x)f]
\end{aligned}$$

hence $(e,1_N)$ is an identity for the monoid $M \wr N$. 7.7

DIVISION

Division

A semigroup S *divides* a semigroup T, denoted $S \preccurlyeq T$, if S is a homomorphic image of a subsemigroup of T. Notice that the divisibility relation $\preccurlyeq$ is reflexive.


Although the divisibility relation is reflexive, most texts use the notation $S \prec T$ instead of $S \preccurlyeq T$.

PROPOSITION 7.8. *The divisibility relation $\preccurlyeq$ is transitive.*

Proof of 7.8. Let S, T, U be semigroups with $S \preccurlyeq T$ and $T \preccurlyeq U$. Then there are subsemigroups T' of T and U' of U and surjective homomorphisms $\varphi : T' \to S$ and $\psi : U' \to T$. Let $U'' = T'\psi^{-1}$. Since T' is a subsemigroup of T, it follows that U'' is a subsemigroup of U' and thus of U. Furthermore, $\psi|_{U'} \circ \varphi : U'' \to S$ is a surjective homomorphism. So $S \preccurlyeq U$. 7.8

The relation of divisibility seems rather 'artificial' here, but it is arises very naturally through the connection between semigroups and finite automata, which we will study in Chapter 9.

PROPOSITION 7.9. *Let S and T be semigroups. Then S and T and their direct product $S \times T$ divide their wreath product $S \wr T$.*

Proof of 7.9. Since S and T are homomorphic images of $S \times T$ under the projection maps $\pi_S : S \times T \to S$ and $\pi_T : S \times T \to T$, we have $S \preccurlyeq S \times T$ and $T \preccurlyeq S \times T$. Since $\preccurlyeq$ is transitive (by Proposition 7.8), it suffices to prove that $S \times T \preccurlyeq S \wr T$.

For each $s \in S$, let $f_s \in S^T$ have all components equal to s. Define a map $\psi : S \times T \to S \wr T$ by $(s,t)\psi = (f_s, t)$. Then

$$\begin{aligned}
&\big((s,t)\psi\big)\big((s',t')\psi\big)\psi\\
&= (f_s,t)(f_{s'},t')\\
&= (f_s\,{}^{t}f_{s'}, tt')\\
&= (f_{ss'}, tt')\\
&\qquad [\text{since } (x)(f_s\,{}^{t}f_{s'}) = (x)f_s(xt)f_{s'} = ss' = (x)f_{ss'} \text{ for all } x \in T]\\
&= (ss', tt')\psi\\
&= \big((s,t)(s',t')\big)\psi.
\end{aligned}$$

So ψ is a homomorphism. Furthermore,

$$\begin{aligned}(s,t)\psi = (s',t')\psi &\Rightarrow (f_s,t) = (f_{s'},t')\\ &\Rightarrow s = s' \wedge t = t'\\ &\Rightarrow (s,t) = (s',t');\end{aligned}$$

thus ψ is injective. Thus $\psi : S \times T \to \operatorname{im}\psi \subseteq S \wr T$ is an isomorphism, and so ψ^{-1} is an surjective homomorphism from the subsemigroup $\operatorname{im}\psi$ of $S \wr T$ to the semigroup $S \times T$. So $S \times T \preccurlyeq S \wr T$. 7.9

PROPOSITION 7.10. *Let M be a monoid and let E be an ideal extension of M by T. Then $E \preccurlyeq T \wr M$.*

Proof of 7.10. By Proposition 1.34, E is a subdirect product of T and M. That is, E is a subsemigroup of $T \times M$ and hence E divides $T \times M$. The result follows from Propositions 7.8 and 7.9. 7.10

PROPOSITION 7.11. *If $S' \preccurlyeq S$ and $T' \preccurlyeq T$, then $S' \wr T' \preccurlyeq S \wr T$.*

Proof of 7.11. The strategy is to prove this in two cases: when S' and T' are subsemigroups of S and T, and when S' and T' are homomorphic images of S and T. The general result follows immediately.

a) Suppose S' and T' are subsemigroups of S and T. Let

$$U = \{(f,t) \in S \wr T : T'f \subseteq S' \wedge t \in T'\}.$$

The immediate aim is to prove that U is a subsemigroup of $S \wr T$. Let $(f_1,t_1),(f_2,t_2) \in U$. So $(f_1,t_1)(f_2,t_2) = (f_1\,{}^{t_1}f_2, t_1t_2)$. First, $t_1t_2 \in T'$ since T' is a subsemigroup of T. Furthermore, for all $x \in T'$,

$$(x)(f_1\,{}^{t_1}f_2) = ((x)f_1)((xt_1)f_2) \in (T'f)(T'f) \subseteq S'$$

since S' is a subsemigroup of S. Hence $(f_1\,{}^{t_1}f_2, t_1t_2) \in U$. Thus U is a subsemigroup of $S \wr T$.

Define $\varphi : U \to S' \wr T'$ by $(f,t)\varphi = (f|_{T'}, t)$. It is clear that φ is a surjective homomorphism and so $S' \wr T' \preccurlyeq S \wr T$.

b) Suppose $\varphi : S \to S'$ and $\psi : T \to T'$ are surjective homomorphisms. Let

$$U = \{(f,t) \in S \wr T : \ker\psi \subseteq \ker(f\varphi)\}. \tag{7.2}$$

As in part a), the first task is to prove that U is a subsemigroup of $S \wr T$. First, note that U is non-empty, because any map $f \in S^T$ with $\ker\psi \subseteq \ker f$ satisfies the condition in (7.2). Now let $(f_1,t_1),(f_2,t_2) \in U$. Let $x, y \in T$ with $x\psi = y\psi$. Then $(x)f_2\varphi = (y)f_2\varphi$ since $\ker\psi \subseteq \ker(f_2\varphi)$. Furthermore, $(xt_1)\psi = (x\psi)(t_1\psi) = (y\psi)(t_1\psi) = (yt_1)\psi$ and so $(xt_1)f_2\varphi = (yt_2)f_2\varphi$ since $\ker\psi \subseteq \ker(f_2\varphi)$. Hence

$$(x)f_1\,{}^{t_1}f_2\varphi = (x)f_1\varphi(xt_1)f_2\varphi = (y)f_1\varphi(yt_1)f_2\varphi = (y)f_1\,{}^{t_1}f_2\varphi.$$

Thus $\ker\psi \subseteq \ker f_1\,{}^{t_1}f_2\varphi$, and so $(f_1,t_1)(f_2,t_2) = (f_1\,{}^{t_1}f_2, t_1t_2) \in U$. So U is a subsemigroup of $S \wr T$.

For any map $f : T \to S$ such that $\ker\psi \subseteq \ker(f\varphi)$, there is a unique map $f' : T' \to S'$ such that $\psi f' = f\varphi$. Define $\vartheta : U \to S' \wr T'$ by $(f,t)\vartheta = (f', t\psi)$. Notice that since ψ is surjective, for any map $f' \in S'^{T'}$ there is a map $f \in S^T$ with $\psi f' = f\varphi$; hence ϑ is surjective. Let $(f_1,t_1),(f_2,t_2) \in U$, then $(f_1,t_1)(f_2,t_2) = (f_1\,{}^{t_1}f_2, t_1t_2)$. Further, $(f_1,t_1)\vartheta(f_2,t_2)\vartheta = (f_1',t_1\psi)(f_2',t_2\psi) = (f_1'\,{}^{t_1\psi}f_2', (t_1t_2)\psi)$. Now

$$\begin{aligned}
&(y\psi)f_1'\,{}^{t_1\psi}f_2' \\
&= (y\psi)f_1'((y\psi)(t_1\psi))f_2' && \text{[by def. of the product and action]}\\
&= (y\psi)f_1'(yt_1)\psi f_2' && \text{[since } \psi \text{ is a homomorphism]}\\
&= (y)f_1\varphi(yt_1)f_2\varphi && \text{[by definition of } f_1' \text{ and } f_2'\text{]}\\
&= \big((y)f_1(yt_1)f_2\big)\varphi && \text{[since } \varphi \text{ is a homomorphism]}\\
&= \big((y)f_1\,{}^{t_1}f_2\big)\varphi, && \text{[by def. of the product and action]}
\end{aligned}$$

and so

$$(f_1\,{}^{t_1}f_2, t_1t_2)\vartheta = (f_1'\,{}^{t_1\psi}f_2', (t_1t_2)\psi). \tag{7.3}$$

Hence

$$\begin{aligned}
&\big((f_1,t_1)(f_2,t_2)\big)\vartheta \\
&= (f_1\,{}^{t_1}f_2, t_1t_2)\vartheta && \text{[multiplication in } S \wr T\text{]}\\
&= (f_1'\,{}^{t_1\psi}f_2', (t_1t_2)\psi) && \text{[by (7.3)]}\\
&= (f_1', t_1\psi)(f_2', t_2\psi) && \text{[factoring in } S' \wr T'\text{]}\\
&= (f_1,t_1)\vartheta(f_2,t_2)\vartheta,
\end{aligned}$$

and thus ϑ is a homomorphism. Therefore $S' \wr T' \preccurlyeq S \wr T$. 7.11

Constant extension

Let S be a semigroup. Let S' be a set in bijection with S under $x \mapsto x'$. Define a multiplication on $S \cup S'$ as follows: multiplication in S is as before (so that S is a subsemigroup of $S \cup S'$), and for all $x, y \in S$,

$$xy' = x'y' = y', \qquad x'y = (xy)'. \tag{7.4}$$

It is easy but tedious to prove that this multiplication is associative (see Exercise 7.9). The set $S \cup S'$ is thus a semigroup, called the *constant extension* of S and denoted $C(S)$.

PROPOSITION 7.12. *If $S \preccurlyeq T$, then $C(S) \preccurlyeq C(T)$.*

Proof of 7.12. If S is a subsemigroup of T, then $C(S)$ is a subsemigroup of $C(T)$.

Suppose S is a homomorphic image of T. Then there exists some surjective homomorphism $\varphi : T \to S$. Define $\widehat{\varphi} : C(T) \to C(S)$ by

$x\widehat{\varphi} = x\varphi$ and $x'\widehat{\varphi} = (x\varphi)'$. Checking the various cases in (7.4) shows that $\widehat{\varphi}$ is a homomorphism. It is clearly surjective. So $C(S)$ is a homomorphic image of $C(T)$.

Hence $S \preccurlyeq T$ implies $C(S) \preccurlyeq C(T)$. 7.12

PROPOSITION 7.13. *Let M be a monoid and S a semigroup. Then $C(S \wr M) \preccurlyeq C(S)^M \wr C(M)$.*


In place of this result, several textbooks claim incorrectly that $C(S \wr M) \preccurlyeq C(S) \wr C(M)$.

Proof of 7.13. Define a map $\psi : C(S \wr M) \to C(S)^M \wr C(M)$ by

$$(f,m)\psi = (f_{\text{ext}}, m), \quad \begin{cases} \text{where } (y)f_{\text{ext}} \in C(S)^M \text{ is defined by} \\ \quad (x)[(y)f_{\text{ext}}] = (xy)f \\ \text{for all } x \in M \text{ and } y \in C(M); \end{cases}$$

$$(f,m)'\psi = (f_{\text{con}}, m'), \quad \begin{cases} \text{where } (y)f_{\text{con}} \in C(S)^M \text{ is defined by} \\ \quad (x)[(y)f_{\text{con}}] = ((x)f)' \\ \text{for all } x \in M \text{ and } y \in C(M). \end{cases}$$

Notice that $(x)[(y)f_{\text{con}}]$ does not depend on y. That is, f_{con} is a constant map from $C(M)$ to $C(S)^M$.


The maps f_{ext} and f_{con} are maps from $C(M)$ to $C(S)^M$. That is, $(y)f_{\text{ext}}$ and $(y)f_{\text{con}}$ are maps from M to $C(S)$ for all $y \in C(M)$. Notice in particular that f_{con} is a constant map from $C(M)$ to $C(S)^M$, but for $y \in C(M)$, the map $(y)f_{\text{con}}$ is in general *not* constant.

We are going to prove that ψ is a monomorphism. Let us first prove that ψ is injective. To begin, observe that $(f,m)\psi = (g,n)'\psi$ implies $(f_{\text{ext}}, m) = (h_{\text{con}}, n')$, which can never happen since $m \in M$ and $n' \in M'$. Thus to prove that ψ is injective, we only have to check that no two distinct elements of $S \wr M$ are mapped to the same element and that no two distinct elements $(S \wr M)'$ are mapped to the same element:

a) Suppose $(f,m)\psi = (g,m)\psi$. Then $f_{\text{ext}} = g_{\text{ext}}$ and $m = n$, and hence $(y)f_{\text{ext}} = (y)f_{\text{ext}}$ for all $y \in M$, or, equivalently, $(xy)f = (xy)g$ for all $x, y \in M$. In particular, putting $x = 1$ shows that $(y)f = (y)g$ for all $y \in M$ and so $f = g$; hence $(f,m) = (g,m)$.

b) Suppose $(f,m)'\psi = (g,n)'\psi$. Then $(f_{\text{con}}, m') = (g_{\text{con}}, n')$ and so $f_{\text{con}} = g_{\text{con}}$ and $m' = n'$. So $((x)f)' = (x)[(y)f_{\text{con}}] = (x)[(y)g_{\text{con}}] = ((x)g)'$, and thus $(x)f = (x)g$ for all $x \in M$. Hence $f = g$ and so $(f,m)' = (g,n)'$.

Therefore ψ is injective.

Next, we have to prove that ψ is a homomorphism. There are four cases to consider, depending on whether each multiplicand lies in $S \wr M$ or $(S \wr M)'$. We explain one case in full here and outline the others; the details are left to Exercise 7.11.

a) Let $(f,m),(g,n) \in S \wr M$. We first have to prove:

$$(f\,{}^m g)_{\mathrm{ext}} = f_{\mathrm{ext}}\,{}^m g_{\mathrm{ext}}. \tag{7.5}$$

Since both sides of (7.5) are maps from $C(M)$ to $C(S)^M$, we must prove that $(y)(f\,{}^m g)_{\mathrm{ext}} = (y) f_{\mathrm{ext}}\,{}^m g_{\mathrm{ext}}$ for all $y \in C(M)$; since both sides of this equality are maps from M to $C(S)$, we must prove that $(x)[(y)(f\,{}^m g)_{\mathrm{ext}}] = (x)[(y) f_{\mathrm{ext}}\,{}^m g_{\mathrm{ext}}]$ for all $x \in M$ and $y \in C(M)$. We proceed as follows:

$$\begin{aligned}
&(x)[(y)(f\,{}^m g)_{\mathrm{ext}}] \\
&= (xy) f\,{}^m g && \text{[by definition of } {}_{\mathrm{ext}}\text{]} \\
&= (xy) f (xym) g && \text{[by def. of the product and action]} \\
&= (x)[(y) f_{\mathrm{ext}}](x)[(y)\,{}^m g_{\mathrm{ext}}] && \text{[by definition of } {}_{\mathrm{ext}}\text{]} \\
&= (x)[(y) f_{\mathrm{ext}} (y)\,{}^m g_{\mathrm{ext}}], && \text{[by multiplication in } C(S)^M\text{]} \\
&= (x)[(y) f_{\mathrm{ext}}\,{}^m g_{\mathrm{ext}}]; && \text{[by multiplication in } (C(S)^M)^{C(M)}\text{]}
\end{aligned}$$

this proves (7.5). Now we have:

$$\begin{aligned}
((f,m)(g,n))\psi &= (f\,{}^m g, mn)\psi \\
&= ((f\,{}^m g)_{\mathrm{ext}}, mn) \\
&= (f_{\mathrm{ext}}\,{}^m g_{\mathrm{ext}}, mn) && \text{[by (7.5)]} \\
&= (f_{\mathrm{ext}}, m)(g_{\mathrm{ext}}, n) \\
&= (f,m)\psi(g,n)\psi.
\end{aligned}$$

b) Let $(f,m)' \in (S \wr M)'$ and $(g,n) \in S \wr M$. By Exercise 7.11,

$$(f\,{}^m g)_{\mathrm{con}} = f_{\mathrm{con}}\,{}^{m'} g_{\mathrm{ext}}. \tag{7.6}$$

Now we have:

$$\begin{aligned}
((f,m)'(g,n))\psi &= (f\,{}^m g, mn)'\psi \\
&= ((f\,{}^m g)_{\mathrm{con}}, (mn)') \\
&= (f_{\mathrm{con}}\,{}^{m'} g_{\mathrm{ext}}, m'n) && \text{[by (7.6)]} \\
&= (f_{\mathrm{con}}, m')(g', n) \\
&= (f,m)'\psi(g,n)\psi.
\end{aligned}$$

c) Let $(f,m) \in S \wr M$ and $(g,n)' \in (S \wr M)'$. By Exercise 7.11,

$$g_{\mathrm{con}} = f_{\mathrm{ext}}\,{}^m g_{\mathrm{con}}. \tag{7.7}$$

Now we have:

$$\begin{aligned}
((f,m)(g,n)')\psi &= (g,n)'\psi \\
&= (g_{\mathrm{con}}, n') \\
&= (f_{\mathrm{ext}}\,{}^m g_{\mathrm{con}}, mn') && \text{[by (7.7)]} \\
&= (f_{\mathrm{ext}}, m)(g_{\mathrm{con}}, n') \\
&= (f,m)\psi(g,n)'\psi.
\end{aligned}$$

d) Let $(f,m)', (g,n)' \in (S \wr M)'$. By Exercise 7.11,

$$g_{\text{con}} = f_{\text{con}}{}^{m} g_{\text{con}}. \tag{7.8}$$

Now we have:

$$\begin{aligned}((f,m)'(g,n)')\psi &= (g,n)'\psi \\ &= (g_{\text{con}}, n') \\ &= (f_{\text{con}}{}^{m} g_{\text{con}}, m'n') && \text{[by (7.8)]} \\ &= (f_{\text{con}}, m')(g_{\text{con}}, n') \\ &= (f,m)'\psi(g,n)'\psi.\end{aligned}$$

Hence ψ is a homomorphism and thus a monomorphism. Therefore ψ^{-1} is a surjective homomorphism from the subsemigroup im ψ of $C(S)^M \wr C(M)$ to $C(S \wr M)$, and so $C(S \wr M) \preccurlyeq C(S)^M \wr C(M)$. 7.13

COROLLARY 7.14. *Let M be a finite monoid and S a semigroup. Then $C(S \wr M)$ divides a wreath product of copies of $C(S)$ and $C(M)$.*

Proof of 7.14. By Proposition 7.13, let M be a monoid and S a semigroup. Then $C(S \wr M) \preccurlyeq C(S)^M \wr C(M)$. But $C(S)^M$ is a direct product of $|M|$ copies of $C(S)$ and so $C(S)^M \wr C(M)$ divides a wreath product of copies of $C(S)$ and $C(M)$ by Propositions 7.9 and 7.11. The result follows by the transitivity of $\preccurlyeq$. 7.14

KROHN–RHODES DECOMPOSITION THEOREM

	1	a	b
1	1	a	b
a	a	a	b
b	b	a	b

TABLE 7.1
Multiplication table of U_3.

Let U_3 be the monoid obtained by adjoining an identity to a two-element right zero semigroup $\{a,b\}$. So U_3 has elements $\{1,a,b\}$ and is multiplication table is as shown in Table 7.1. Notice that $x^2 = x$ for all $x \in U_3$, and so U_3 is aperiodic.

The Krohn–Rhodes theorem is often stated in the form 'every finite semigroup divides a wreath product of finite groups and finite aperiodic semigroups'. We will prove a stronger form by showing that every finite semigroup divides a wreath product of its own subgroups and copies of U_3. We first of all note that it suffices to prove the theorem for monoids since $S \preccurlyeq S^1$. The proof is by induction on the number of elements in the monoid. The core of the induction is Lemma 7.16, which shows that a monoid S is either a group, a left simple semigroup with an identity adjoined, monogenic, or can be decomposed as $S = L \cup T$, where L is a left ideal and T is a submonoid and L^1 and T have fewer elements than S. The theorem is trivial for groups, and we will prove it for left simple semigroups with identities adjoined (Lemma 7.18) and for monogenic

semigroups (Lemma 7.19); these cases form the base of the induction. The fourth possibility, of decomposition as $S = L \cup T$, supplies the induction step. The reader may wish to look ahead to Figure 7.1 on page 147 to keep track of the roles of the various lemmata.

We need the following auxiliary result before we prove Lemma 7.16:

LEMMA 7.15. *Let S be a finite semigroup. Then at least one of the following is true:*

a) *S is trivial;*

b) *S is left simple;*

c) *S is monogenic;*

d) *$S = L \cup T$, where L is a proper left ideal of S and T is a proper subsemigroup of S.*

Proof of 7.15. Suppose that none of the properties a), b), or c) is true; we aim to prove d). Since S is not left simple, it contains proper left ideals. Since it is finite, it has a maximal proper left ideal K.

Let $x \in S \setminus K$. Then $K \cup S^1x$ is a left ideal that strictly contains K. Since K is maximal, $K \cup S^1x = S$. If $S^1x \neq S$, then let $L = K$ and $T = S^1x$ and the proof is complete.

So assume $S^1x = S$. Then $S == Sx \cup \{x\} \subseteq Sx \cup \langle x \rangle$. If $S \neq Sx$, then let $L = Sx$ and $T = \langle x \rangle$ and the proof is complete since $T \neq S$ because S is not monogenic.

So assume $S = Sx$. Let $M = \{ y \in S : yx \in K \}$. Then M non-empty (since $Mx = K$) and is a left ideal of S. Furthermore, it is a proper left ideal because K is a proper left ideal and $Sx = S$. If $M \not\subseteq K$, then $M \cup K$ is a left ideal of S strictly containing the maximal left ideal K and so $M \cup K = S$; set $L = M$ and $T = K$ and the proof is complete.

So assume $M \subseteq K$; that is,

$$yx \in K \Rightarrow y \in K. \tag{7.9}$$

Repeat the reasoning above for all $x \in S \setminus K$. Either some such x allows us to complete the proof, or (7.9) holds for all $x \in S \setminus K$. In the former case, the proof is finished. In the latter case, take the contrapositive to see that $y \in S \setminus K \Rightarrow yx \in S \setminus K$ for all $x \in S \setminus K$. Therefore $S \setminus K$ is a subsemigroup. So let $L = K$ and $T = S \setminus K$; the proof is complete. 7.15

LEMMA 7.16. *Let S be a finite monoid. Then at least one of the following is true:*

a) *S is a group;*

b) *S is a left simple with an identity adjoined;*

c) *S is monogenic;*

d) *$S = L \cup T$, where L is a left ideal of S and T is a submonoid of S, and L^1 and T both have fewer elements than S.*

Proof of 7.16. Suppose that none of the properties a), b), and c) is true; we aim to prove d). Let G be the group of units of S. Consider two cases:

a) G is trivial. Then $S \setminus G = S \setminus \{1\}$ is an ideal by Proposition 7.1 and thus a subsemigroup of S. Since S is not left simple with an identity adjoined, we know that $S \setminus \{1\}$ is not left simple. Apply Lemma 7.15 to $S \setminus \{1\}$ to see that $S \setminus \{1\} = L \cup Q$, where L is a proper left ideal of $S \setminus \{1\}$ and Q is a proper subsemigroup of $S \setminus \{1\}$. Since $L \neq S \setminus \{1\}$, we know that $L \cup \{1\} \neq S$. Let $T = Q \cup \{1\}$; then T is a proper submonoid of S and $S = L \cup T$.

b) G is non-trivial. Then let $L = S \setminus G$ and let $T = G$. Then $S = L \cup T$. Since G is non-trivial, $L \cup \{1\} \neq S$, and since S is not a group, $T \neq S$.

In both cases, $S = L \cup T$, where L is a left ideal of S and T is a submonoid of S. Furthermore, in both cases L^1 and T both have fewer elements than the original monoid S. 7.16

Now we turn to proving the cases forming the base of the induction; that is, monoids consisting of a left simple semigroup with an identity adjoined, and monogenic monoids. To prove the former case in Lemma 7.18, we will need the following lemma, which essentially shows that the theorem holds for left zero semigroups:

LEMMA 7.17. *Every finite left zero semigroup divides a wreath product of copies of U_3.*

Proof of 7.17. Let L_n be a left zero semigroup with n elements. The strategy is to proceed by induction and show that $L_n^1 \preccurlyeq L_{n-1}^1 \wr L_1^1$. The base of the induction is proved by observing that the semigroup $L_1^1 = \{0, 1\}$ is a homomorphic image of U_3.

For each $x \in L_{n-1}^1$, define $f_x : L_1^1 \to L_{n-1}^1$ by $(1)f_x = x$ and $(0)f_x = 1$. Let

$$K = \{ (f_x, 0) : x \in L_{n-1}^1 \}.$$

Furthermore, $(f_x, 0)(f_y, 0) = (f_x\, {}^0f_y, 0) = (f_x, 0)$, since for all $z \in L_1^1$,

$$(z)f_x\, {}^0f_y = (z)f_x(z0)f_y = (z)f_x(0)f_y = (z)f_x1 = (z)f_x.$$

Hence K is a left zero subsemigroup of $L_{n-1}^1 \wr L_1^1$. Notice that $n = |L_{n-1}^1| = |K|$. Furthermore, the wreath product $L_{n-1}^1 \wr L_1^1$ is a monoid by Proposition 7.7. Therefore $K \cup \{1\}$ is a subsemigroup of $L_{n-1}^1 \wr L_1^1$ isomorphic to L_n^1. Hence $L_n^1 \preccurlyeq L_{n-1}^1 \wr L_1^1$.

By Proposition 7.11, L_n^1 divides a wreath product of n copies of U_3; thus L_n also divides a wreath product of n copies of U_3 by the transitivity of divisibility. 7.17

LEMMA 7.18. *Let S be a finite left simple semigroup. Then S^1 divides the wreath product of a subgroup of S and copies of U_3.*

Proof of 7.18. Since S is finite, it contains an idempotent. Thus by Theorem 4.19, we see that S is isomorphic to $Z \times G$, where Z is a left zero semigroup and G is a subgroup of S. By Lemma 7.17, Z divides a wreath product of copies of U_3. So by Propositions 7.9 and 7.11, $Z \times G$ divides a wreath product of G and copies of U_3. 7.18

We now turn to proving the remaining case in the base of the induction, namely monogenic monoids:

LEMMA 7.19. *Let S be a finite monogenic monoid. Then S divides the wreath product of a subgroup of S and copies of U_3.*

Proof of 7.19. The monogenic monoid $S = \{1, x, \dots, x^k, \dots, x^{k+m-1}\}$ (with $x^{k+m} = x^k$) is an ideal extension of the subgroup $G = \{x^k, \dots, x^{k+m-1}\}$ by the monogenic monoid $C_k = \{1, x, \dots, x^k\}$ (with $x^k = x^k + 1$).

We proceed by induction on k and show that C_k divides a subsemigroup of $C_{k-1} \wr C_1$. The base case of the induction is proven by observing that $C_1 = \{1, x\}$ (with $x^2 = x$) divides U_3, since it is isomorphic to the subsemigroup $\{1, a\}$ of U_3.

For $i \in \mathbb{N}$, define $f_i : C_1 \to C_{k-1}$ by $(1)f_i = x^{i-1}$ and $(0)f_i = x^i$. Let

$$U = \{1\} \cup \{(f_i, 0) : i \in \mathbb{N}\} \subseteq C_{k-1} \wr C_1.$$

Let $(f_i, 0)(f_j, 0) \in U$. Then $(f_i, 0)(f_j, 0) = ({f_i}^0 f_j, 0) = (f_{i+j}, 0)$ since

$$\begin{aligned}(0){f_i}^0 f_j &= (0)f_i(0)f_j = x^i x^j = x^{i+j} = (0)f_{i+j},\\ (1){f_i}^0 f_j &= (1)f_i(0)f_j = x^{i-1}x^j = (1)f_{i+j}.\end{aligned}$$

Hence U is a submonoid of $C_{k-1} \wr C_1$. In particular, $(f_i, 0) = (f_1, 0)^i_)$ for all $i \in \mathbb{N}$, and so U is the monogenic submonoid of $C_{k-1} \wr C_1$ generated by $(f_1, 0)$. Finally, note that

$$\begin{aligned}(f_1, 0)^{k+1} &= (f_{k+1}, 0) = (f_k, 0) = (f_1, 0)^k,\\ (f_1, 0)^k &= (f_k, 0) \neq (f_{k-1}, 0) = (f_1, 0)^{k-1};\end{aligned}$$

and so

$$U = \{1, (f_1, 0), (f_1, 0)^2, \dots, (f_1, 0)^k\}.$$

Hence U is isomorphic to C_k, and therefore $C_k \preccurlyeq C_{k-1} \wr C_1$.

Thus every C_k divides a wreath product of U_3 by Propositions 7.9 and 7.11. So S, being an ideal extension of G and C_k, divides a wreath product of G and copies of U_3 by Proposition 7.10. 7.19

Finally, we are ready to being proving the induction step, in the case where the monoid has been decomposed as the union of a left ideal and a subsemigroup. We require the following four lemmata, and then we can quickly prove the theorem.

LEMMA 7.20. *Let S be a semigroup and suppose $S = L \cup T$, where L is a left ideal of S and T is a subsemigroup of S. Then $S \preccurlyeq L^1 \wr C(T^1)$.*

Proof of 7.20. Let $i : C(T^1) \to L^1$ be the constant map defined by $(t)i = (t')i = 1$ for all $t \in T^1$. For each $x \in L$, let $f_x : C(T^1) \to L^1$ be the right translation defined by $(t)f_x = (t')f_x = tx$ for all $t \in T^1$. Notice that $tx \in L$ since L is a left ideal.

Let

$$V = \{ (i,t) : t \in T^1 \} \cup \{ (f_x, t') : x \in L, t \in T^1 \}.$$

We aim to show V is a subsemigroup of $L^1 \wr C(T^1)$ and that S is a homomorphic image of V. We have four cases to consider:

a) Let $(i,t),(i,u) \in V$. Then $(i,t)(i,u) = (i\,{}^t i, tu) = (i,tu)$ since

$$(s)i\,{}^t i = (s)i(st)i = 1 = (s)i,$$
$$(s')i\,{}^t i = (s')i(s't)i = 1 = (s')i$$

for all $s \in T^1$.

b) Let $(i,t),(f_y,u') \in V$. Then $(i,t)(f_y,u') = (i\,{}^t f_y, tu') = (f_{ty}, u')$, since

$$(s)i\,{}^t f_y = (s)i(st)f_y = 1sty = (s)f_{ty},$$
$$(s')i\,{}^t f_y = (s')i(s't)f_y = 1s'ty = (s')f_{ty}$$

for all $s \in T^1$.

c) Let $(f_x,t'),(i,u) \in V$. Then $(f_x,t')(i,u) = ({f_x}^{t'} i, t'u) = (f_x, (tu)')$ since

$$(s){f_x}^{t'} i = (s)f_x(st')i = (s)f_x 1 = (s)f_x,$$
$$(s'){f_x}^{t'} i = (s')f_x(s't')i = (s')f_x 1 = (s')f_x$$

for all $s \in T^1$.

d) Let $(f_x,t'),(f_y,u') \in V$. Then

$$(f_x,t')(f_y,u') = ({f_x}^{t'} f_y, t'u') = (f_{xty}, u'),$$

since

$$(s){f_x}^{t'} f_y = (s)f_x(st')f_y = (s)f_x(t')f_y = sxty = (s)f_{xty},$$
$$(s'){f_x}^{t'} f_y = (s')f_x(s't')f_y = (s')f_x(t')f_y = sxty = (s')f_{xty}$$

for all $s \in T^1$.

Hence V is a subsemigroup of $L^1 \wr C(T^1)$. Define a map $\varphi : V \to S$ by $(i,t)\varphi = t$ and $(f_x, t')\varphi = xt$. Then, using the four cases above,

$$\begin{aligned}
\big((i,t)(i,u)\big)\varphi &= (i,tu)\varphi = tu = (i,t)\varphi(i,u)\varphi,\\
\big((i,t)(f_y,u')\big)\varphi &= (f_{ty},u')\varphi = tyu = (i,t)\varphi(f_y,u')\varphi,\\
\big((f_x,t')(i,u)\big)\varphi &= (f_x,(tu)')\varphi = xtu = (f_x,t')\varphi(i,u)\varphi,\\
\big((f_x,t')(f_y,u')\big)\varphi &= (f_{xty},u')\varphi = xtyu = (f_x,t')\varphi(f_y,u')\varphi;
\end{aligned}$$

hence φ is a homomorphism. Finally, note that $T \subseteq \operatorname{im}\varphi$ since $(i,t)\varphi = t$ for all $t \in T$ and $L \subseteq \operatorname{im}\varphi$ since $(f_x,1)\varphi = x$ for all $x \in L$. Hence $S = L \cup T = \operatorname{im}\varphi$ and so φ is a surjective homomorphism. Thus $S \preccurlyeq L^1 \wr C(T^1)$. 7.20

The following result is essentially a more precise version of Lemma 7.20 that holds when we decompose a monoid into the union of its group of units and the set of remaining elements:

LEMMA 7.21. *Let S be a monoid and let G be its group of units. Then $I = S \setminus G$ is an ideal of S and $S \preccurlyeq I^1 \wr G$.*

Proof of 7.21. First, notice that $S \setminus G$ is an ideal by Proposition 7.1. For each $x \in I^1$, define a map $f_x : G \to I^1$ by $(g)f_x = gxg^{-1}$ for all $g \in G$. Notice that $(g)f_1 = gg^{-1} = 1$ for all $g \in G$. Let $V = \{(f_x, g) : x \in I^1, g \in G\}$.

Let $(f_x, g), (f_y, h) \in V$. Then for any $k \in G$,

$$\begin{aligned}
(k)f_x\,{}^g f_y &= (k)f_x(kg)f_y\\
&= kxk^{-1}kgyg^{-1}k^{-1}\\
&= k(xgyg^{-1})k^{-1}\\
&= (k)f_{xgyg^{-1}}.
\end{aligned}$$

Therefore

$$(f_x,g)(f_y,h) = (f_x\,{}^g f_y, gh) = (f_{xgyg^{-1}}, gh). \tag{7.10}$$

Notice that $xgyg^{-1} \in I^1$ since x is in I^1 and I is an ideal. Thus V is a subsemigroup of $I^1 \wr G$.

Define a map $\varphi : V \to S$ by $(f_x, g)\varphi = xg$. This map φ is well-defined since $f_x = f_y \Rightarrow (1)f_x = (1)f_y \Rightarrow x = y$. Furthermore,

$$\begin{aligned}
\big((f_x,g)(f_y,h)\big)\varphi &= (f_{xgyg^{-1}}, gh)\varphi && \text{[by (7.10)]}\\
&= xgyg^{-1}gh\\
&= xgyh\\
&= (f_x,g)\varphi(f_y,h)\varphi.
\end{aligned}$$

So φ is a homomorphism. Finally, $G \subseteq \operatorname{im}\varphi$ since $(f_1, g)\varphi = g$ for all $g \in G$ and $I^1 \subseteq \operatorname{im}\varphi$ since $(f_x, 1)\varphi = x$ for all $x \in S$. So φ is surjective and so $S \preccurlyeq I^1 \wr G$. 7.21

The following lemma shows that the result holds for right zero semigroups, but we only use it to prove the next lemma.

LEMMA 7.22. *Every finite right zero semigroup, and every finite right zero semigroup with an identity adjoined, divides a wreath product of copies of U_3.*

Proof of 7.22. Let R_k denote the right zero semigroup with k elements. Notice that R_k is a subsemigroup of R_ℓ and R_k^1 is a submonoid of R_ℓ^1 for $k \leqslant \ell$. The direct product of n copies of U_3 contains subsemigroups isomorphic to R_{2^n} and $R_{2^n}^1$, so for any k the direct product of sufficiently many copies of U_3 contains R_k and R_k^1. So R_k and R_k^1 divide a wreath product of copies of U_3 by Proposition 7.9. 7.22

LEMMA 7.23. *Let S be a finite semigroup. If S divides a wreath product of groups and copies of U_3, then $C(S)$ divides a wreath product of copies of those same groups and copies of U_3.*

Proof of 7.23. Supppose that S divides a wreath product of groups G_i and copies of U_3. By Propositions 7.11, 7.12, and 7.14, $C(S)$ divides a wreath product of the monoids $C(G_i)$ and copies of $C(U_3)$. Now, the semigroup $C(U_3)$ is a right zero semigroup with an identity adjoined, which divides a wreath product of copies of U_3 by Lemma 7.22. The group of units of $C(G_i)$ is G_i and $L = C(G_i) \setminus G_i$ is an ideal of $C(G_i)$ by Proposition 7.1. Furthermore, L is a right zero semigroup by (7.4). So $C(G_i)$ divides $L^1 \wr G_i$ by Lemma 7.21. Since L^1 is a right zero semigroup with an identity adjoined, it divides the wreath product of copies of U_3 by Lemma 7.22. So $L^1 \wr G_i$ divides a wreath product of G_i and copies of U_3 by Proposition 7.11. So S divides a wreath product of the groups G_i and copies of U_3. 7.23

Finally, using these lemmata, we can proof the Krohn–Rhodes Theorem. To keep track of the roles of the various lemmata, see Figure 7.1.

Krohn–Rhodes theorem

KROHN–RHODES THEOREM 7.24. *Let S be a finite semigroup. Then S divides a wreath product of subgroups of S and copies of U_3.*

Proof of 7.24. Let S be a semigroup; we will show that S divides a wreath product of its subgroups and copies of U_3. Since $S \preccurlyeq S^1$, we can assume S is a monoid.

The strategy is induction on the number of elements in S. The base case of the induction is when S has one element. In this case, S is trivial, and so S is a group and the result holds immediately.

So assume the result holds for all monoids with fewer elements than S. As already noted, the result clearly holds if S is a group and in particular if S is trivial. It also holds by Lemma 7.18 if S is a left simple semigroup with an identity adjoined, and if S is monogenic by Lemma 7.19.

So assume S is not trivial, not a group, not a left simple semigroup with an identity adjoined, and not monogenic. By Lemma 7.16, $S = L \cup T$,

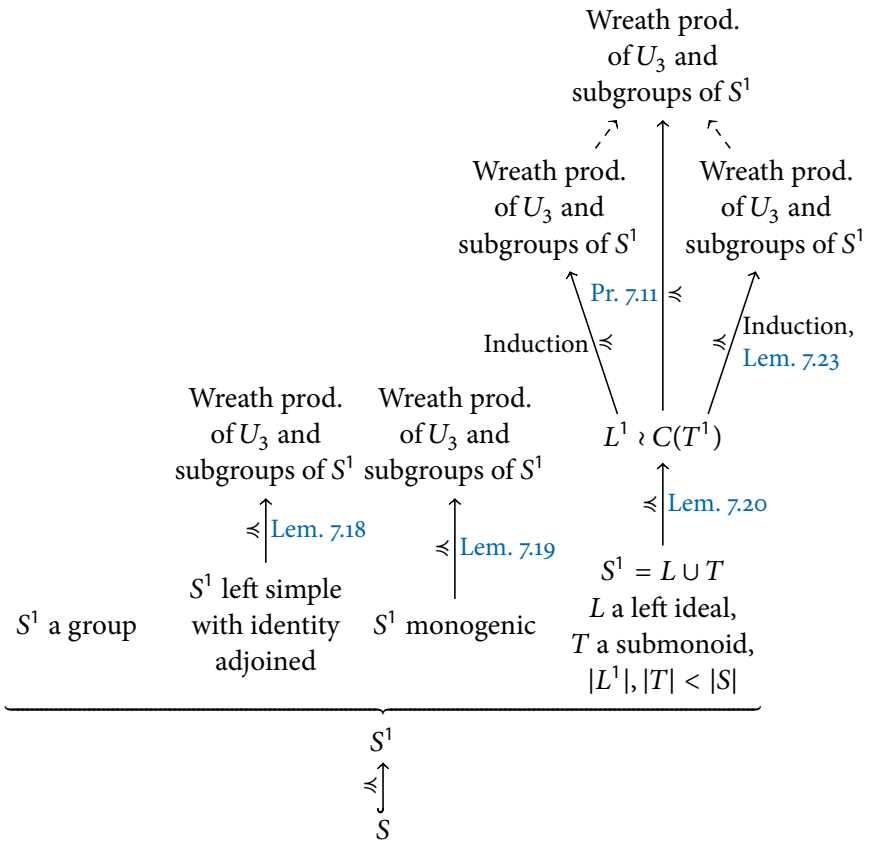


FIGURE 7.1
Diagram showing how the various lemmata are used to prove the Krohn–Rhodes theorem.

where L is a left ideal and T is a submonoid of S and both L^1 and T have fewer elements than S. So by the induction hypothesis, L^1 divides a wreath product of subgroups of L^1 (which are also subgroups of S) and copies of U_3, and similarly T divides a wreath product of subgroups of T (which are also subgroups of S) and copies of U_3. By Lemma 7.23, $C(T)$ also divides a wreath product of subgroups of S and copies of U_3. By Proposition 7.11, S thus divides a wreath product of subgroups of S and copies of U_3.

Thus, by induction, the result holds for all monoids S. 7.24

EXERCISES

[See pages 237–241 for the solutions.]

7.1 Let M be a finite monoid. Prove that M is a group if and only if $MxM = M$ for all $x \in M$.

7.2 Let S be a finite semigroup. Let J_x be a nontrivial $\mathcal{J}$-class of S. Prove that there is a regular $\mathcal{J}$-class J_y such that $J_x \leqslant J_y$.

*7.3 a) Prove that a finite nilsemigroup is nilpotent.

b) Give an example of an infinite nilsemigroup that is not nilpotent.

7.4 Let S and S' be finite semigroups and let $\varphi : S \to S'$ be a surjective homomorphism.

a) Let J be a $\mathcal{J}$-class of S. Prove that there is a $\mathcal{J}$-class J' of S' such that $J\varphi \subseteq J'$.

b) Let J' be a $\mathcal{J}$-class of S'. Prove that there is a $\mathcal{J}$-class J of S such that $J\varphi \subseteq J'$. If J is minimal such that $J\varphi \subseteq J'$, then $J\varphi = J'$.

✶7.5 Prove that if S is a finite semigroup in which $\mathcal{H}$ is the equality relation and $T \preccurlyeq S$, then in T the relation $\mathcal{H}$ is also the equality relation. Give an example to show that this is may not be true when S is infinite.

✶7.6 Prove that the multiplication defined for semidirect products (7.1) is associative.

7.7 Prove that if M and N are groups, $M \wr N$ is a group.

7.8 Suppose that S and T are cancellative semigroups. Must $S \wr T$ be cancellative?

✶7.9 Prove that the product defined by (7.4) for the constant extension is associative.

7.10 Let M be a non-trivial monoid. For each $x \in M$, let $\rho_x \in \mathcal{T}_M$ and $\tau_x \in \mathcal{T}_M$ be defined by $y\rho_x = yx$ and $y\tau_x = x$. Prove that $C(M)$ is isomorphic to the subset $\{\rho_x, \tau_x : x \in M\}$ of $\mathcal{T}_M$.

✶7.11 Using a technique similar to the proof of (7.5), prove (7.6), (7.7), (7.8)

NOTES

The Krohn–Rhodes theorem was first stated and proved for automata in Krohn & Rhodes, 'Algebraic theory of machines I'. ◆ The proof in this chapter is due to Lallement, *Semigroups and Combinatorial Applications*, and incorporates the correction published in Lallement, 'Augmentations and wreath products of monoids'. ◆ Rhodes & Steinberg, *The q-theory of Finite Semigroups* is the most comprehensive monograph on finite semigroup theory, but is decidedly non-elementary.

Varieties & pseudovarieties

8

> ‘ Variety of opinion is necessary for objective knowledge. And a method that encourages variety is also the only method that is compatible with a humanitarian outlook. ’
>
> — Paul Feyeraband, *Against Method*, § 3.

The aim of this chapter is to introduce varieties and pseudovarieties of semigroups and monoids. These are classes of that are well-behaved and can, in particular, be defined using sets of equations. For instance, the class of all commutative semigroups forms a variety, and the class of all *finite* commutative semigroups forms a pseudovariety, and both are defined by the equation $xy = yx$. We will formalize these notions later in the chapter, and we will see how varieties and pseudovarieties can be defined and manipulated in different ways.

Pseudovarieties are important in the study of finite semigroups for the following reason: there are many finite semigroups of a given size, but most of them are boring. Of the 3 684 030 417 non-isomorphic semigroups with 8 elements, 3 661 522 792 of them are nilpotent semigroups S satisfying $S^3 = \{0\}$. (Essentially, the reason there are so many of these semigroups is that any multiplication in which all products of length 3 are equal to 0 is trivially associative.) Pseudovarieties allow us to isolate the more interesting classes.

The concepts of varieties and pseudovarieties are actually broader than semigroups: varieties make sense for any type of algebraic structure, and pseudovarieties make sense for any type of finite algebraic structure. Therefore we will begin by discussing varieties in terms of universal algebra.

VARIETIES

Algebras and operations

An *algebra* is a set S equipped with some operations $\{ f_i : i \in I \}$. An *operation* f_i on S is simply a map $f_i : S^{f_i\alpha} \to S$ for some $f_i\alpha \in \mathbb{N} \cup \{0\}$. This $f_i\alpha$ is called the *arity* of f_i. For instance, if S is a semigroup, the multiplication operation $\circ$ is a map $\circ : S^2 \to S$ and so has arity 2. If S is a inverse semigroup, the inverse operation $^{-1}$ is a map

Arity of an operation

$^{-1} : S \to S$ and so has arity 1. If S is a monoid, we can view the identity element 1_S as an operation, or as a map $1_S : S^0 \to S$ (where $S^0 = \varnothing$); the operation 1_S has arity 0. Operations of arity 1 are called *unary*; operations of arity 2 are called *binary*; operations of arity 0 are called *constants*. Notice that we have a map $\alpha : \{ f_i : i \in I \} \to \mathbb{N} \cup \{0\}$.

Types

A *type* $\mathcal{T}$ of an algebra is a set of operations symbols $\{ f_i : i \in I \}$ and a map $\alpha : \{ f_i : i \in I \} \to \mathbb{N} \cup \{0\}$ determining the arity of each operations. We can write the type simply by listing the pairs in the map α (viewed as a set). A semigroup has type $\{(\circ, 2)\}$, a monoid has type $\{(\circ, 2), (1, 0)\}$ (the identity operation 1 is constant), and a lattice has type $\{(\sqcap, 2), (\sqcup, 2)\}$. An algebra of type $\mathcal{T}$ is called a *$\mathcal{T}$-algebra*.

Notice that some structures can be viewed as algebras in more than one way, and thus have more than one type. Let G be a group. Then G, viewed as a group, is a $\{(\circ, 2), (1_G, 0), (^{-1}, 1)\}$-algebra; G, viewed as a monoid, is a $\{(\circ, 2), (1_G, 0)\}$-algebra; G, viewed as a semigroup, is a $\{(\circ, 2)\}$-algebra.

Strictly speaking, we should distinguish a symbol f_i from the operation f_i: for instance, we use the same symbol $\circ$ to refer to the different multiplications in different semigroups. We will want to use the same symbol to discuss operations of the same arity in different structures.

Let $\mathcal{T} = \{ (f_i, f_i\alpha) : i \in I \}$ be a type. We are now going to give the definition of subalgebras, homomorphisms, congruences, and direct products of $\mathcal{T}$-algebras. These definitions are straightforward generalizations of the definitions for semigroups.

Subalgebras

Let S be a $\mathcal{T}$-algebra. A subset S' of S is a *subalgebra* of S if S' is closed under all the operations in $\mathcal{T}$: that is, for each $i \in I$, we have

$$x_1, \ldots, x_{f_i\alpha} \in S' \Rightarrow (x_1, \ldots, x_{f_i\alpha}) f_i \in S'. \tag{8.1}$$

In particular, this means that $f_i \in S'$ whenever $f_i\alpha = 0$. Let $X \subseteq S$. The subalgebra generated by X is defined to be the intersection of all subalgebras that contain X. It is easy to prove (cf. Proposition 1.11) that the subalgebra generated by X consists of all elements that can be obtained by starting from X and applying the operations f_i.

Homomorphisms

Let S and T be $\mathcal{T}$-algebras. Then $\varphi : S \to T$ is a *homomorphism* if for each $i \in I$, we have

$$\big((x_1, \ldots, x_{f_i\alpha}) f_i\big)\varphi = (x_1\varphi, \ldots, x_{f_i\alpha}\varphi) f_i. \tag{8.2}$$

(Notice that on the left-hand side of (8.2), f_i is an operation on S, while on the right-hand side, it is an operation on T.) An injective homomorphism is a *monomorphism*, and a bijective homomorphism is an *isomorphism*. If $\varphi : S \to T$ is a surjective homomorphism, T is a *homomorphic image* of S.

Congruences

Let S be a $\mathcal{T}$-algebra. A binary relation ρ on S is a *congruence* if for

each $i \in I$,

$$(\forall x_1, y_1, \ldots, x_{f_i\alpha}, y_{f_i\alpha})$$
$$(x_1 \ \rho \ y_1 \wedge \ldots \wedge x_{f_i\alpha} \ \rho \ x_{f_i\alpha}$$
$$\Rightarrow (x_1, \ldots, x_{f_i\alpha})f_i \ \rho \ (y_1, \ldots, y_{f_i\alpha})f_i).$$

Direct products

Let $\mathcal{S} = \{ S_j : j \in J \}$ be a collection of $\mathcal{T}$-algebras. The *direct product* of the $\mathcal{T}$-algebras in $\mathcal{S}$ is their cartesian product $\prod_{j\in J} S_j$ with the operations performed componentwise:

$$(j)(s_1, \ldots, s_{f_i\alpha})f_i = \big((j)s_1, \ldots, (j)s_{f_i}\big)f_i.$$

Free $\mathcal{T}$-algebras

Let A be a non-empty set and let $F_{\mathcal{T}}(A)$ be the smallest set of all formal expressions (that is, words) over $A \cup \{ f_i : i \in I \} \cup \{(\} \cup \{)\} \cup \{,\}$ satisfying the following two conditions:

$$A \subseteq F_{\mathcal{T}}(A);$$
$$u_1, \ldots, u_{f_i\alpha} \in F_{\mathcal{T}}(A) \Rightarrow (u_1, \ldots, u_{f_i\alpha})f_i \in F_{\mathcal{T}}(A).$$

For instance, if $\mathcal{T}$ is $\{(f, 2), (\ ', 1)\}$ and $A = \{a, b, c\}$, then the words $(a, (((c, b)f)', c)f)f$ and $(((b)', ((b, a)f, (c)')f)f$ are elements of $F_{\mathcal{T}}(A)$. The set $F_{\mathcal{T}}(A)$ is obviously a $\mathcal{T}$-algebra and is called the *free $\mathcal{T}$-algebra* or *absolutely free $\mathcal{T}$-algebra*. Notice that $F_{\mathcal{T}}(A)$ is generated by A.

Let $\iota : A \to F_{\mathcal{T}}(A)$ be the inclusion map. For any $\mathcal{T}$-algebra S and map $\varphi : A \to S$, there is a unique extension of φ to a homomorphism $\widehat{\varphi} : F_{\mathcal{T}}(A) \to S$. That is, $\iota\widehat{\varphi} = \varphi$, or, equivalently, the following diagram commutes:

$$\begin{array}{ccc} A & \xhookrightarrow{\iota} & F_{\mathcal{T}}(A) \\ & \searrow^{\varphi} & \downarrow{\scriptstyle \widehat{\varphi}} \\ & & S \end{array} \tag{8.3}$$

This property is reminiscent of some definitions we have already seen: free semigroups and monoids (see pages 38 f.), free inverse semigroups and monoids (see page 107), and free commutative semigroups (see page 123), and we shall say more about it later.

$\mathbb{H}$, $\mathbb{S}$, $\mathbb{P}$

Let $\mathcal{X}$ be a non-empty class of $\mathcal{T}$-algebras. Let $\mathbb{H}\mathcal{X}$ denote the class of all $\mathcal{T}$-algebras that are homomorphic images of the algebras in $\mathcal{X}$. Let $\mathbb{S}\mathcal{X}$ denote the class of all $\mathcal{T}$-algebras that are subalgebras of algebras in $\mathcal{X}$. Let $\mathbb{P}\mathcal{X}$ denote the class of all $\mathcal{T}$-algebras that are direct products of the $\mathcal{T}$-algebras in $\mathcal{X}$. That is,

$$\mathbb{H}\mathcal{X} = \{ S : (\exists T \in \mathcal{X})(S \text{ is a homomorphic image of } T) \};$$
$$\mathbb{S}\mathcal{X} = \{ S : (\exists T \in \mathcal{X})(S \text{ is a subalgebra of } T) \};$$
$$\mathbb{P}\mathcal{X} = \{ S : (\exists \{ T_i : i \in I \} \subseteq \mathcal{X})\big(S = \textstyle\prod_{i\in I} T_i\big) \}.$$

Thus $\mathbb{H}$, $\mathbb{S}$, and $\mathbb{P}$ are unary operators on classes of algebras. Notice that $\mathcal{X}$ is contained in $\mathbb{H}\mathcal{X}$, $\mathbb{S}\mathcal{X}$, and $\mathbb{P}\mathcal{X}$.

Variety

A non-empty class of $\mathcal{T}$-algebras is a *variety* of $\mathcal{T}$-algebras if it is closed under the operations $\mathbb{H}$, $\mathbb{S}$, and $\mathbb{P}$. That is, $\mathcal{X}$ is a variety if $\mathbb{H}\mathcal{X} \cup \mathbb{S}\mathcal{X} \cup \mathbb{P}\mathcal{X} \subseteq \mathcal{X}$.

EXAMPLE 8.1. a) Let $\mathsf{1}$ be the class containing only the trivial semigroup $E = \{e\}$. Then $\mathsf{1}$ is a variety, since the only subsemigroup of E is E itself, the only homomorphic image of E is E itself, and any direct product of copies of E is isomorphic to E.

b) Let S be the class of all semigroups (viewed as $\{(\circ, 2)\}$-algebras). Any homomorphic image of a semigroup is itself a semigroup, so S is closed under $\mathbb{H}$. Subalgebras are subsemigroups, and so S is closed under $\mathbb{S}$. A direct product of semigroups is a semigroup, so S is closed under $\mathbb{P}$. Therefore S is a variety.

c) Let M be the class of all monoids (viewed as $\{(\circ, 2), (1, 0)\}$-algebras). A homomorphic image of a monoid is again a monoid, so M is closed under $\mathbb{H}$. Subalgebras are submonoids because the subalgebra must be closed under the 'operation' 1: that is, they must contain the constant 1. So M is closed under $\mathbb{S}$. A direct product of monoids is itself a monoid. Therefore M is a variety.

d) Let Com be the class of all commutative semigroups (viewed, like members of S, as $\{(\circ, 2)\}$-algebras). Since any subalgebra or homomorphic image of a commutative semigroup is itself a commutative semigroup, and a direct product of commutative semigroups is commutative, Com is a variety.

e) Let G be the class of all groups, viewed as $\{(\circ, 2), (1_G, 0), (^{-1}, 1)\}$-algebras. Then subalgebras are closed under multiplication and taking inverses; thus subalgebras are subgroups. Since any subalgebra or homomorphic image of a group is also a group, and any direct product of groups is also a group, G is a variety.

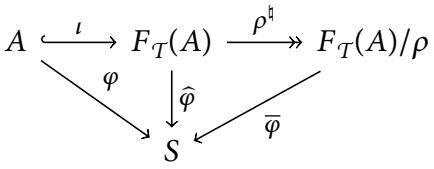
Notice that the class of all groups $\mathcal{G}$ viewed as $\{(\circ, 2)\}$-algebras is *not* a variety, because in this case subalgebras are subsemigroups and so $\mathcal{G}$ is not closed under taking subalgebras; for example, $\mathcal{G}$ contains $\mathbb{Z}$ but not its subsemigroup $\mathbb{N}$.

f) Let Inv be the class of inverse semigroups, viewed as $\{(\circ, 2), (^{-1}, 1)\}$-algebras. Then Inv is a variety.

Let V be a variety of $\mathcal{T}$-algebras and let A be a non-empty set. Let $S \in \mathsf{V}$. Let $\varphi \in S^A$. We know there is a unique extension of φ to a homomorphism $\widehat{\varphi} : F_{\mathcal{T}}(A) \to S$. Now, $\operatorname{im} \widehat{\varphi}$ is a subalgebra of S and so $\operatorname{im} \widehat{\varphi} \in \mathsf{V}$ since V is closed under forming subalgebras. Let

$$\rho = \bigcap \{ \ker \widehat{\varphi} : \varphi \in S^A, S \in \mathsf{V} \}.$$

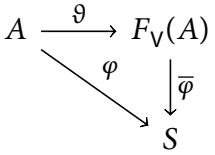
Then ρ, being an intersection of congruences on $F_{\mathcal{T}}(A)$, is also a congruence. Furthermore, $\rho \subseteq \ker \widehat{\varphi}$ for any $\varphi \in S^A$. Hence for each $S \in \mathsf{V}$ and $\varphi \in S^A$, there exists a unique homomorphism $\overline{\varphi} : F_{\mathcal{T}}(A)/\rho \to S$ such that $\rho^\natural \overline{\varphi} = \widehat{\varphi}$. Thus $\overline{\varphi} : F_{\mathcal{T}}(A)/\rho \to S$ is the unique homomorphism such that $\varphi = \iota\widehat{\varphi} = \iota\rho^\natural\overline{\varphi}$, and the following diagram commutes:



By a result for $\mathcal{T}$-algebras analogous to Proposition 1.33, $F_{\mathcal{T}}(A)/\rho$ is a subdirect product of $\{ F_{\mathcal{T}}(A)/\ker\widehat{\varphi} : \varphi \in S^A, S \in \mathsf{V} \}$. Now, each algebra $F_{\mathcal{T}}(A)/\ker\widehat{\varphi}$ is a subalgebra of an element of V and therefore is itself a member of V (since $\mathbb{S}\mathsf{V} = \mathsf{V}$). Hence $F_{\mathcal{T}}(A)/\rho \in \mathbb{S}\mathbb{P}\mathsf{V} = \mathsf{V}$.

The $\mathcal{T}$-algebra $F_{\mathcal{T}}(A)/\rho$ is called the V-*free algebra*, and is denoted $F_{\mathsf{V}}(A)$. Notice that there is a map $\vartheta : A \to F_{\mathsf{V}}(A)$ given by $x\vartheta = x\iota\rho^\natural = [x]_\rho$ such that the *universal property* holds: for any $S \in \mathsf{V}$ and map $\varphi : A \to S$, there is a unique homomorphism $\overline{\varphi} : F_{\mathsf{V}}(A) \to S$ such that $\vartheta\overline{\varphi} = \varphi$, or, in diagrammatic terms:

V-free algebras



(8.4)

Notice that if V is the variety of all $\mathcal{T}$-algebras, $F_{\mathsf{V}}(A) = F_{\mathcal{T}}(A)$ and we recover diagram (8.3).

Let us apply this definition to some concrete varieties. Let V be the variety of all semigroups S. Then the definition of a S-free algebra coincides with the definition of a free semigroup, and the diagram (8.4) becomes identical to the second diagram in (2.1). Since $F_{\mathsf{S}}(A) \in \mathsf{S}$, we see that $F_{\mathsf{S}}(A) \simeq A^+$ by Proposition 2.1.

Similarly, if we apply the definition to the variety of inverse semigroups Inv, the diagram (8.4) becomes identical to the second diagram in (5.10) and we see that $F_{\mathsf{Inv}}(A) \simeq \mathrm{FInvS}(A)$ by Proposition 5.15. With the variety of commutative semigroups Com, the diagram (8.4) becomes identical to the second diagram in (6.1) and we see that $F_{\mathsf{Com}}(A) \simeq \mathrm{FCommS}(A)$ by Proposition 6.3.

Thus for any variety V of $\mathcal{T}$-algebras we have (for each set A) a $\mathcal{T}$-algebra $F_{\mathsf{V}}(A) \in \mathsf{V}$ and a map $\vartheta : A \to F_{\mathsf{V}}(A)$ with the universal property. This indicates that varieties have some of the nice properties of the classes of semigroups, inverse semigroups, and commutative semigroups. But varieties also have another very useful property: they are precisely those collections of algebras that can be defined using sets of equations called laws.

Laws

For $\mathcal{T}$-algebras, a *law* over an alphabet A is a pair of elements u and v of $F_{\mathcal{T}}(A)$, normally written as a formal equality $u = v$. A $\mathcal{T}$-algebra S *satisfies* the law $u = v$ if, for every map $\varphi : A \to S$, we have $u\widehat{\varphi} = v\widehat{\varphi}$ (where $\widehat{\varphi}$ is the homomorphism in diagram (8.3)). Informally, S satisfies $u = v$ if we every possible substitution of elements of S for letters of A in the words u and v gives elements that are equal. For instance, commutative semigroups, viewed as $\{(\circ, 2)\}$-algebras, satisfy the law $x \circ y = y \circ x$. Semigroups of idempotents satisfy the law $x \circ x = x$. All semigroups satisfy the law $x \circ (y \circ z) = (x \circ y) \circ z$, and all monoids satisfy the laws $x \circ 1 = x$ and $1 \circ x = x$.



A law over A is sometimes called an *identity* or *identical relation* over A, but we will avoid this potentially confusing terminology.

Equational classes

Let $\mathcal{E}$ be a class of $\mathcal{T}$-algebras. Suppose there is a set L of laws over an alphabet A such that $S \in \mathcal{E}$ if and only if S satisfies every law in L. Then $\mathcal{E}$ is the *equational class* defined by L.

Birkhoff's theorem

BIRKHOFF'S THEOREM 8.2. *Let $\mathcal{T}$ be a type. Then a class of $\mathcal{T}$-algebras is a variety if and only if it is an equational class.*

Proof of 8.2. Part 1. Suppose $\mathcal{X}$ is an equational class. Then there is a set of laws L over an alphabet A such that $S \in \mathcal{X}$ if and only if S satisfies every law in L. To prove that $\mathcal{X}$ is a variety, we must show that it is closed under $\mathbb{H}$, $\mathbb{S}$, and $\mathbb{P}$.

Let $S \in \mathcal{X}$, and let T be a $\mathcal{T}$-algebra and $\psi : S \to T$ a surjective homomorphism. Let $u = v$ be a law in L. Let $\varphi : A \to T$ be a map. Define a map $\vartheta : A \to S$ by letting $a\vartheta \in S$ be such that $a\vartheta\psi = a\varphi$ (such an $a\vartheta$ exists because ψ is surjective). Notice that $\widehat{\vartheta}\psi$ and $\widehat{\varphi}$ are homomorphisms from $F_{\mathcal{T}}(A)$ to S extending $\vartheta\psi = \varphi$ and so, by the uniqueness of such homomorphisms, $\widehat{\vartheta}\psi = \widehat{\varphi}$. Since S satisfies L, we have $u\widehat{\vartheta} = v\widehat{\vartheta}$; hence $u\widehat{\varphi} = u\widehat{\vartheta}\psi = v\widehat{\vartheta}\psi = v\widehat{\varphi}$. So T satisfies $u = v$. Hence T satisfies every law in L and so $T \in \mathcal{X}$. Thus $\mathcal{X}$ is closed under $\mathbb{H}$.

Let $S \in \mathcal{X}$ and let T be a subalgebra of S. Let $u = v$ be a law in L. Then if $\varphi : A \to T$, then φ is also a map from A to S and so $u\widehat{\varphi} = v\widehat{\varphi}$ since S satisfies $u = v$. Hence T also satisfies $u = v$. So T satisfies every law in L and so $T \in \mathcal{X}$. Thus $\mathcal{X}$ is closed under $\mathbb{S}$.

Let $\{S_j : j \in J\} \subseteq \mathcal{X}$ and suppose T is the direct product of $\{S_j : j \in J\}$. Let $u = v$ be a law in L. Let $\varphi : A \to T$ be a map. For each $j \in J$, let $\varphi_j = \varphi\pi_j$, where $\pi_j : T \to S_j$ is the projection homomorphism. So φ_j is a map from A to S_j, and S_j satisfies $u = v$, and thus $u\widehat{\varphi}_j = v\widehat{\varphi}_j$. The map $\psi : F_{\mathcal{T}}(A) \to T$ with $(j)(x\psi) = x\widehat{\varphi}_j$ is a homomorphism extending φ, so, by the uniqueness condition, $\psi = \widehat{\varphi}$. Since $u\widehat{\varphi}_j = v\widehat{\varphi}_j$ for each j, we have $u\widehat{\varphi} = u\psi = v\psi = v\widehat{\varphi}$; hence T satisfies $u = v$. So T satisfies every law in L and so $T \in \mathcal{X}$. Thus $\mathcal{X}$ is closed under $\mathbb{P}$.

So $\mathcal{X}$ is closed under $\mathbb{H}$, $\mathbb{S}$, and $\mathbb{P}$, and so is a variety.

Part 2. Suppose now that V is a variety. Let A be an infinite alphabet.

Recall that $F_{\mathsf{V}}(A) = F_{\mathcal{T}}(A)/\rho$, where

$$\rho = \bigcap \{ \ker \widehat{\varphi} : \varphi \in S^A, S \in \mathsf{V} \}.$$

We aim to show that V is the equational class defined by ρ, viewing the set of pairs $\rho \subseteq F_{\mathcal{T}}(A) \times F_{\mathcal{T}}(A)$ as a set of laws.

Let $S \in \mathsf{V}$. Let $(u, v) \in \rho$; notice that $u, v \in F_{\mathcal{T}}(A)$. Then $(u, v) \in \ker \widehat{\varphi}$ and thus $u\widehat{\varphi} = v\widehat{\varphi}$ for any $\varphi \in S^A$. So S satisfies the law $u = v$. Thus every $S \in \mathsf{V}$ satisfies the law $u = v$ for any $(u, v) \in \rho$.

Conversely, suppose that S satisfies the law $u = v$ for every $(u, v) \in \rho$. Let B be an alphabet with cardinality greater than or equal to both S and A. Let $F_{\mathsf{V}}(B)$ be the V-free algebra generated by B; then $F_{\mathsf{V}}(B) = F_{\mathcal{T}}(B)/\pi$ for some congruence π on $F_{\mathcal{T}}(B)$.

Since B has cardinality greater than or equal to S, there is a surjective homomorphism $\psi : F_{\mathcal{T}}(B) \to S$. We are going to prove that $\pi \subseteq \ker \psi$, which will imply that we have a well-defined surjective homomorphism $\vartheta : F_{\mathsf{V}}(B) \to S$ with $[x]_\pi \vartheta = x\psi$, which will in turn imply $S \in \mathsf{V}$.

So let $(u, v) \in \pi$. Let B_0 be the subset of B containing the letters that appear in u or v; notice that B_0 is finite. Let A_0 be a finite subset of A such that there is a bijection $\xi_0 : A_0 \to B_0$. Since B has cardinality greater than or equal to A, there is an injection $\xi : A \to B$ extending ξ_0. Since ξ is injective, there is a right inverse $\eta : B \to A$ of ξ (that is, $\xi\eta = \mathrm{id}_A$). Then ξ extends to a monomorphism $\widehat{\xi} : F_{\mathcal{T}}(A) \to F_{\mathcal{T}}(B)$, and η extends to a homomorphism $\widehat{\eta} : F_{\mathcal{T}}(B) \to F_{\mathcal{T}}(A)$. Since $\widehat{\xi}$ is injective, there are uniquely determined $u_0, v_0 \in F_{\mathcal{T}}(A)$ such that $u_0\widehat{\xi} = u$ and $v_0\widehat{\xi} = v$. Notice that $u\eta = u_0$ and $v\eta = v_0$.

Consider the map $\eta\rho^\natural : F_{\mathcal{T}}(B) \to F_{\mathsf{V}}(A)$. Since $F_{\mathsf{V}}(B)$ lies in the variety V, we must have $\pi \subseteq \eta\rho^\natural$. In particular, $u\eta\rho^\natural = v\eta\rho^\natural$ and so $(u_0, v_0) \in \rho$. Thus S satisfies the law $u_0 = v_0$. Therefore, since $\eta\psi : F_{\mathcal{T}}(B) \to S$ is a homomorphism, $u_0\eta\psi = v_0\eta\psi$ and so $u\psi = v\psi$.

Hence $\pi \subseteq \ker \psi$ and so we have a well-defined homomorphism $\vartheta : F_{\mathsf{V}}(B) \to S$ with $[x]_\pi \vartheta = x\psi$. Therefore S is a homomorphic image of $F_{\mathsf{V}}(B) \in \mathsf{V}$ and so lies in the variety V.

Thus we have proved that $S \in \mathsf{V}$ if and only if S satisfies every law $u = v$ in ρ. Therefore V is an equational class. 8.2

Finitely based variety

Theorem 8.2 shows that every variety can be defined by a set of laws. However, in general an infinite set of laws is required. This is true even for varieties of semigroups. However, in some cases, a finite set of laws suffice. Such varieties are said to be *finitely based*.

Let $\mathcal{T}$ be the type $\{(\circ, 2)\}$. The variety consisting of all semigroups S is defined by the law $x \circ (y \circ z) = (x \circ y) \circ z$. We use this type when working with varieties of semigroups, and we will always implicitly assume this law and write xy for $x \circ y$. Examples are summarized in Table 8.1.

Let $\mathcal{T}$ be the type $\{(\circ, 2), (1, 0)\}$. The variety consisting of all monoids M is defined by the laws $x(yz) = (xy)z$, $1x = x$, and $x1 = x$. When

Variety	Symbol	Defining laws
Semigroups	S	—
Null semigroups	Z	$xy = zt$
Left zero semigroups	LZ	$xy = x$
Right zero semigroups	RZ	$xy = y$

TABLE 8.1
Varieties of semigroups. The law $x(yz) = (xy)z$ is implicitly assumed.

Variety	Symbol	Defining laws
Monoids	M	—
Trivial monoid	1	$x = 1$
Commutative monoids	Com	$xy = yx$
Semilattices with identities	Sl	$\begin{cases} x^2 = x, \\ xy = yx \end{cases}$

TABLE 8.2
Varieties of monoids, viewed as $\{(\circ, 2), (1, 0)\}$-algebras. The laws $x(yz) = (xy)z$, $x1 = x$, and $1x = x$ are implicitly assumed.

working with varieties of monoids we will use this type and implicitly assume these laws. Examples are summarized in Table 8.2.

Finally let $\mathcal{T}$ be the type $\{(\circ, 2), (^{-1}, 1)\}$. We will use this type when working with semigroups equipped with an inverse operation $^{-1}$, such as regular and inverse semigroup. In this context, we will assume the laws $(xy)z = x(yz)$, $xx^{-1}x = x$ and $(x^{-1})^{-1} = x$. Examples are summarized in Table 8.3.

Variety generated by $\mathcal{X}$

Another way to define a variety of $\mathcal{T}$-algebras is to use a specified set of $\mathcal{T}$-algebras to generate a variety. Let $\mathcal{X}$ be a set of $\mathcal{T}$-algebras. The intersection of all varieties of $\mathcal{T}$-algebras containing $\mathcal{X}$ is itself a variety, called the *variety of $\mathcal{T}$-algebras generated by $\mathcal{X}$*, or simply the *variety generated by $\mathcal{X}$*. It is easy to prove that the variety generated by $\mathcal{X}$ consists of all $\mathcal{X}$-algebras that can be obtained from $\mathcal{X}$ by repeatedly forming subsemigroups, homomorphic images, and direct products. That is, the variety generated by $\mathcal{X}$ is

$$\{\, \mathbb{O}_1\mathbb{O}_2\cdots\mathbb{O}_n\mathcal{X} : n \in \mathbb{N}, \mathbb{O}_i \in \{\mathbb{H}, \mathbb{S}, \mathbb{P}\}\,\}. \tag{8.5}$$

LEMMA 8.3. *For any non-empty class of $\mathcal{T}$-algebras $\mathcal{X}$, we have*

$$\mathbb{S}\mathbb{H}\mathcal{X} \subseteq \mathbb{H}\mathbb{S}\mathcal{X};$$
$$\mathbb{P}\mathbb{H}\mathcal{X} \subseteq \mathbb{H}\mathbb{P}\mathcal{X};$$
$$\mathbb{P}\mathbb{S}\mathcal{X} \subseteq \mathbb{S}\mathbb{P}\mathcal{X}.$$

Proof of 8.3. Let $S \in \mathbb{S}\mathbb{H}\mathcal{X}$. Then there is $\mathcal{T}$-algebra $T \in \mathcal{X}$ and a surjective homomorphism $\varphi : T \to U$ such that S is a subalgebra of U. Let $T' = S\varphi^{-1} = \{\, t \in T : t\varphi \in S\,\}$. Then T' is a subalgebra of T and $\varphi|_{T'} : T' \to S$ is a surjective homomorphism. So $S \in \mathbb{H}\mathbb{S}\mathcal{X}$.

Let $S \in \mathbb{P}\mathbb{H}\mathcal{X}$. Then there is a collection of $\mathcal{T}$-algebras $\{\, T_i : i \in I\,\} \subseteq \mathcal{X}$ and a collection of surjective homomorphisms $\Phi = \{\, \varphi_i : T_i \to U_i : i \in I\,\}$ such that $S = \prod_{i\in I} U_i$. Define a homomorphism $\psi : \prod_{i\in I} T_i \to S$ by

Variety	Symbol	Defining laws
Completely regular sgrps	CR	$xx^{-1} = x^{-1}x$
Inverse semigroups	Inv	$\begin{cases}(xy)^{-1} = y^{-1}x^{-1},\\ xx^{-1}yy^{-1} = yy^{-1}xx^{-1}\end{cases}$
Clifford semigroups	Cl	$\begin{cases}xx^{-1} = x^{-1}x,\\ xx^{-1}yy^{-1} = yy^{-1}xx^{-1}\end{cases}$
Groups	G	$xx^{-1} = yy^{-1}$

TABLE 8.3
Varieties of semigroups with a unary operation $^{-1}$. The laws $x(yz) = (xy)z$, $xx^{-1}x = x$, and $(x^{-1})^{-1} = x$ are implicitly assumed.

$(i)(x\psi) = ((i)x)\varphi_i$. Then ψ is surjective since each φ_i is surjective. So $S \in \mathbb{HIP}\mathcal{X}$.

Let $S \in \mathbb{IPS}\mathcal{X}$. Then there is a collection of $\mathcal{T}$-algebras $\{T_i : i \in I\} \subseteq \mathcal{X}$ and a subalgebras U_i of T_i such that $S = \prod_{i\in I} U_i$. Then S is a subalgebra of $\prod_{i\in I} T_i$. So $S \in \mathbb{SIP}\mathcal{X}$. 8.3

As an immediate consequence of Lemma 8.3 and (8.5), and the fact that the operators $\mathbb{H}$, $\mathbb{S}$, and $\mathbb{P}$ are idempotent, we obtain the following result:

PROPOSITION 8.4. *Let $\mathcal{X}$ be a class of $\mathcal{T}$-algebras. The variety generated by $\mathcal{X}$ is $\mathbb{HSIP}\mathcal{X}$.* 8.4

PSEUDOVARIETIES

Varieties are not useful for studying and classifying finite algebras, for the simple reason that every non-trivial variety contains infinite algebras: if a variety contains an algebra S with two elements, then it contains the direct product of infinitely many copies of S, which is of course infinite.

Clearly, if we take a class $\mathcal{X}$ of finite $\mathcal{T}$-algebras, then $\mathbb{H}\mathcal{X}$ and $\mathbb{S}\mathcal{X}$ also contain only finite $\mathcal{T}$-algebras. The problem, therefore, is the operator $\mathbb{P}$. To modify the notion of variety in order to study finite algebras, we therefore introduce a new operator on classes of $\mathcal{T}$-algebras.

Let $\mathbb{P}_{\text{fin}}\mathcal{X}$ denote the class of all $\mathcal{T}$-algebras that are finitary direct products of the algebras in $\mathcal{X}$. That is, $\mathbb{P}_{\text{fin}}$

$$\mathbb{P}_{\text{fin}}\mathcal{X} = \{S : (\exists\{T_1, \dots, T_n\} \subseteq \mathcal{X})(S = T_1 \times T_2 \times \dots \times T_n)\}.$$

A non-empty class of finite $\mathcal{T}$-algebras is a *pseudovariety* of $\mathcal{T}$-algebras if it is closed under the operations $\mathbb{H}$, $\mathbb{S}$, and $\mathbb{P}_{\text{fin}}$. That is, $\mathcal{X}$ is a pseudovariety if $\mathbb{H}\mathcal{X} \cup \mathbb{S}\mathcal{X} \cup \mathbb{P}_{\text{fin}}\mathcal{X} \subseteq \mathcal{X}$. Pseudovariety

EXAMPLE 8.5. a) Let 1 be the class containing only the trivial semigroup (or monoid) $E = \{e\}$. Then 1 is a pseudovariety both when

we view E as a $\{(\circ, 2)\}$-algebra and when we view E as a $\{(\circ, 2), (1, 0)\}$-algebra, since the only subalgebra of E is E itself, the only homomorphic image of E is E itself, and any finitary direct product of copies of E is isomorphic to E.

b) Let S be the class of all finite semigroups. Then S is a pseudovariety.

c) Let M be the class of all finite monoids (viewed as $\{(\circ, 2), (1, 0)\}$-algebras. Then S is a pseudovariety.

d) Let Com be the class of all finite commutative monoids (viewed as $\{(\circ, 2), (1, 0)\}$-algebras). Then Com is a pseudovariety.

e) Let G be the class of all finite groups, which we view as $\{(\circ, 2), (1_G, 0), (^{-1}, 1)\}$-algebras; then G is a pseudovariety.

In contrast with varieties, that the class of all finite groups viewed as $\{(\circ, 2)\}$-algebras *is* a pseudovariety, because in this case subalgebras are subgroups (since for elements x of a group with n elements, $x^n = 1$ and $x^{-1} = x^{n-1}$).

f) Let Inv be the class of all finite inverse semigroups, which we view as $\{(\circ, 2), (^{-1}, 1)\}$-algebras. Then Inv is a pseudovariety.

g) Let N be the class of all finite nilpotent semigroups. Then N is a pseudovariety. (See Exercise 8.2(a).)

h) Let A be the class of all finite aperiodic monoids, viewed as $\{(\circ, 2), (1, 0)\}$-algebras. Then A is a pseudovariety. Notice that $\mathsf{N} \subseteq \mathsf{A}$.

Notice that we are using the same symbols for certain varieties and pseudovarieties: for instance, Com is used to denote both the variety of commutative monoids and the pseudovariety of finite commutative monoids. This will not cause confusion, because from now on we will only use them to denote pseudovarieties.

Pseudovariety generated by $\mathcal{X}$

Just as with varieties, we have the idea of generating a pseudovariety of finite $\mathcal{T}$-algebras. Let $\mathcal{X}$ be a set of finite $\mathcal{T}$-algebras. The intersection of all pseudovarieties of $\mathcal{T}$-algebras containing $\mathcal{X}$ is itself a pseudovariety, called the *pseudovariety of finite $\mathcal{T}$-algebras generated by $\mathcal{X}$*, or simply the *pseudovariety generated by $\mathcal{X}$*, and is denoted $\mathsf{V}_\mathcal{T}(\mathcal{X})$ It is easy to prove that $\mathsf{V}_\mathcal{T}(\mathcal{X})$ consists of all (necessarily finite) $\mathcal{X}$-algebras that can be obtained from $\mathcal{X}$ by repeatedly forming subalgebras, homomorphic images, and finitary direct products. That is,

$$\mathsf{V}_\mathcal{T}(\mathcal{X}) = \{\, \mathbb{O}_1\mathbb{O}_2\cdots\mathbb{O}_n\mathcal{X} : n \in \mathbb{N}, \mathbb{O}_i \in \{\mathbb{H}, \mathbb{S}, \mathbb{P}_{\text{fin}}\} \,\}. \tag{8.6}$$

We have the following analogue of Proposition 8.4:

PROPOSITION 8.6. *Let $\mathcal{X}$ be a class of finite $\mathcal{T}$-algebras. Then* $\mathsf{V}_\mathcal{T}(\mathcal{X}) = \mathbb{H}\mathbb{S}\mathbb{P}_{\text{fin}}\mathcal{X}$.

Proof of 8.6. For any non-empty class of finite $\mathcal{T}$-algebras $\mathcal{X}$, we have

$$\mathbb{S}\mathbb{H}\mathcal{X} \subseteq \mathbb{H}\mathbb{S}\mathcal{X};$$
$$\mathbb{P}_{\text{fin}}\mathbb{H}\mathcal{X} \subseteq \mathbb{H}\mathbb{P}_{\text{fin}}\mathcal{X};$$
$$\mathbb{P}_{\text{fin}}\mathbb{S}\mathcal{X} \subseteq \mathbb{S}\mathbb{P}_{\text{fin}}\mathcal{X};$$

to see this, follow the reasoning in the proof of Lemma 8.3, restricting the index sets I in the direct products to be finite. The result follows immediately. 8.6

Join and meet of pseudovarieties

Let V and W be pseudovarieties of $\mathcal{T}$-algebras. The class of pseudovarieties of $\mathcal{T}$-algebras is ordered by the usual inclusion order $\subseteq$. Then it is easy to see that

$$\mathsf{V} \sqcup \mathsf{W} = \mathsf{V}_{\mathcal{T}}(\mathsf{V} \cup \mathsf{W}),$$

and, since $\mathsf{V} \cap \mathsf{W}$ is a variety,

$$\mathsf{V} \sqcap \mathsf{W} = \mathsf{V} \cap \mathsf{W}.$$

So the class of $\mathcal{T}$-algebras is a lattice. Furthermore, if we consider only subpseudovarieties of a fixed pseudovariety V (such as S or M), then the class of such subpseudovarieties forms a sublattice.

PSEUDOVARIETIES OF SEMIGROUPS AND MONOIDS

From this point onwards, we will consider only pseudovarieties of semigroups and pseudovarieties of monoids. These pseudovarieties have different types: pseudovarieties of semigroups have type $\mathcal{S} = \{(\circ, 2)\}$ and pseudovarieties of monoids have type $\mathcal{M} = \{(\circ, 2), (1, 0)\}$.

In pseudovarieties of semigroups, the homomorphisms are the usual semigroup homomorphisms and the subalgebras are subsemigroups. In pseudovarieties of monoids, the homomorphisms are *monoid* homomorphisms, and the subalgebras are submonoids *that contain the identity of the original monoid*. In previous chapters, by 'submonoid' we meant 'any subsemigroup that forms a monoid'. But such a submonoid may not be a subalgebra: For example, let $S = \{1, 0\}$ be the two-element semilattice with $1 > 0$. Then S is a monoid with identity 1, and contains the submonoid $T = \{0\}$. However, T is not an $\mathcal{M}$-subalgebra of S, because an $\mathcal{M}$-subalgebra must include the constant 1. For brevity, we call a submonoid that contains the identity of the original monoid an *$\mathcal{M}$-submonoid*.

$\mathcal{S}$-pseudovarieties, $\mathcal{M}$-pseudovarieties

We will use the term *$\mathcal{S}$-pseudovarieties* for pseudovarieties of semigroups, and *$\mathcal{M}$-pseudovarieties* for pseudovarieties of monoids. The reasoning for the two types often runs in parallel, but there are important differences.

Notice that $\mathcal{S}$-pseudovarieties are closed under division: if V is an $\mathcal{S}$-pseudovariety, $T \in \mathsf{V}$, and $S \preccurlyeq T$, then by definition there is an surjective homomorphism $\varphi : T' \to S$, where T' is a subsemigroup of T; hence $S \in \mathbb{H}\mathbb{S}\mathsf{V} = \mathsf{V}$. In fact, $\mathcal{M}$-pseudovarieties are also closed under division, as a consequence of the following result:

Division in $\mathcal{M}$-pseudovarieties

PROPOSITION 8.7. *Let S be a semigroup and M a monoid and suppose $S \preccurlyeq M$. Then there is an $\mathcal{M}$-submonoid M' of M and a surjective monoid homomorphism $\varphi : M' \to S^1$. Consequently, if N is a monoid, then $N \preccurlyeq M$ if and only if there is an $\mathcal{M}$-submonoid M' of M and a surjective monoid homomorphism $\varphi : M' \to N$.*

Proof of 8.7. Suppose $S \preccurlyeq M$. Then there is a subsemigroup T of M and a surjective homomorphism $\psi : T \to S$. Let $M' = T \cup \{1_M\}$ and extend ψ to a monoid homomorphism $\varphi : M' \to S$ by defining $1_M\varphi = 1_{S^1}$ and $x\varphi = x\psi$ for all $x \in T$. (Notice that φ is well-defined, since if $1_M \in T$, then $(z\psi)(1_M\psi) = (z1_M)\psi = z\psi$ and $(1_M\psi)(z\psi) = (1_M z)\psi = z\psi$ and so S is a monoid with identity $1_M\psi$ because ψ is surjective.) 8.7

We now introduce two operators that allow us to connect $\mathcal{S}$-pseudovarieties of semigroups and $\mathcal{M}$-pseudovarieties of monoids.

V_{Sg} For any $\mathcal{M}$-pseudovariety of monoids V, let

$$\mathsf{V}_{\mathrm{Sg}} = \mathsf{V}_{\mathcal{S}}(\mathsf{V}).$$

So to obtain V_{Sg} from V we simply treat the monoids in V as semigroups, form all finite direct products, then all subsemigroups, and then all (semigroup) homomorphic images. From Proposition 8.7, we see that

$$S \in \mathsf{V}_{\mathrm{Sg}} \Leftrightarrow S^1 \in \mathsf{V}. \tag{8.7}$$

$\mathsf{V}_{\mathrm{Mon}}$ For any $\mathcal{S}$-pseudovariety of semigroups, let

$$\mathsf{V}_{\mathrm{Mon}} = \{ S \in \mathsf{V} : S \text{ is a monoid} \}.$$

That is $\mathsf{V}_{\mathrm{Mon}}$ consists of the monoids that, when viewed as semigroups, belong to V. It is easy to see that $\mathsf{V}_{\mathrm{Mon}}$ is an $\mathcal{M}$-pseudovariety of monoids.

LEMMA 8.8. *For any $\mathcal{M}$-pseudovariety of monoids V, we have $(\mathsf{V}_{\mathrm{Sg}})_{\mathrm{Mon}} = \mathsf{V}$.*

Proof of 8.8. Let S be a finite monoid. Then

$$\begin{aligned}
& S \in (\mathsf{V}_{\mathrm{Sg}})_{\mathrm{Mon}} \\
\Leftrightarrow\ & S \text{ is a monoid that belongs to } \mathsf{V}_{\mathrm{Sg}} \\
\Leftrightarrow\ & S^1 \in \mathsf{V} && \text{[by (8.7)]} \\
\Leftrightarrow\ & S \in \mathsf{V}. && \text{[since } S = S^1\text{]} \quad 8.8
\end{aligned}$$

PROPOSITION 8.9. *The operator* $_{\mathrm{Sg}}$ *is an embedding of the lattice of* $\mathcal{M}$*-pseudovarieties of monoids into the lattice of of the* $\mathcal{S}$*-pseudovarieties of semigroups.*

Proof of 8.9. It is immediate from (8.7) that $_{\mathrm{Sg}}$ is a lattice homomorphism. By Lemma 8.8,

$$\mathsf{V}_{\mathrm{Sg}} = \mathsf{W}_{\mathrm{Sg}} \Rightarrow (\mathsf{V}_{\mathrm{Sg}})_{\mathrm{Mon}} = (\mathsf{W}_{\mathrm{Sg}})_{\mathrm{Mon}} \Rightarrow \mathsf{V} = \mathsf{W},$$

so $_{\mathrm{Sg}}$ is injective. 8.9

An $\mathcal{S}$-pseudovariety of semigroups W is *monoidal* if $\mathsf{W} = \mathsf{V}_{\mathrm{Sg}}$ for some $\mathcal{M}$-pseudovariety of monoids. Monoidal pseudovariety

EXAMPLE 8.10. a) The $\mathcal{S}$-pseudovariety of all finite semigroups S is monoidal, because $\mathsf{S} = \mathsf{M}_{\mathrm{Sg}}$ by (8.7), where M is the $\mathcal{M}$-pseudovariety of all finite monoids.

b) The $\mathcal{S}$-pseudovariety of all finite nilpotent semigroups N is not monoidal. To see this, suppose, with the aim of obtaining a contradiction, that $\mathsf{N} = \mathsf{V}_{\mathrm{Sg}}$ for some $\mathcal{M}$-pseudovariety of all finite monoids V. Then $\mathsf{N}_{\mathrm{Mon}} = \mathsf{V}$ by (8.8). Let $M \in \mathsf{N}$ be a monoid. Then $1_M^n = 0_M$ for some $n \in \mathbb{N}$, since M is nilpotent, which implies that M is trivial. Hence $\mathsf{N}_{\mathrm{Mon}} = 1$, and so $\mathsf{N} = \mathsf{V}_{\mathrm{Sg}} = (\mathsf{N}_{\mathrm{Mon}})_{\mathrm{Sg}} = 1_{\mathrm{Sg}} = 1$, which is a contradiction.

FREE OBJECTS FOR PSEUDOVARIETIES

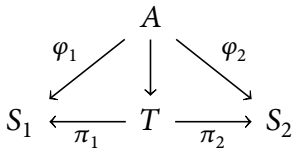
If we want to follow the same path for pseudovarieties as for varieties, our next step should be to construct a 'free V-semigroup' for each $\mathcal{S}$-pseudovariety V and a 'free W-monoid' for each $\mathcal{M}$-pseudovariety W, and then to devise an analogue of laws and prove an analogue of Birkhoff's theorem. However, this is much more difficult for pseudovarieties than for varieties. We will outline the problems and describe the solution in this section and the next two sections. To simplify the explanation, we will only discuss $\mathcal{S}$-pseudovarieties, but every result and construction in these sections has a parallel for $\mathcal{M}$-pseudovarieties, replacing semigroups with monoids, homomorphisms with monoid homomorphisms, and subsemigroups with $\mathcal{M}$-submonoids as appropriate.

The basic problem in finding free objects for pseudovarieties is very simple: free objects are usually infinite, and members of a pseudovariety are always finite. Consider the $\mathcal{S}$-pseudovariety N of finite nilpotent semigroups. For any finite alphabet A and $n \in \mathbb{N}$, let $I_n = \{w \in A^+ : |w| \geqslant n\}$. Then A^+/I_n is a nilpotent semigroup; thus $A^+/I_n \in \mathsf{N}$. The semigroup A^+/I_n contains at least n elements (and indeed contains $|A|^n$ elements if

$|A| \geqslant 2$). Thus, by taking n to be arbitrarily large, we see that N contains arbitrarily large A-generated semigroups. Since an A-generated free object for N must map surjectively to each of these semigroups, it is clear that no semigroup in N is free.

If we try to approach the idea of a free object through laws, we encounter another problem. It is clear that A^+/I_n satisfies no law in at most $|A|$ variables where the two sides of the law have length less than n. So if we try to base our free objects on laws, all $\mathcal{S}$-pseudovarieties containing N will have the same free object.

Let us look at free objects from another direction. The idea is that a free A-generated object for a class $\mathcal{X}$ should be just general enough to be more general than any A-generated object in $\mathcal{X}$. Suppose we take two semigroups S_1 and S_2 in an $\mathcal{S}$-pseudovariety V. Let $\varphi_1 : A \to S_1$ and $\varphi_2 : A \to S_2$ be functions such that $\operatorname{im}\varphi_1$ generates S_1 and $\operatorname{im}\varphi_2$ generates S_2. Let T be the subsemigroup of $S_1 \times S_2$ generated by $\{(a\varphi_1, a\varphi_2) : a \in A\}$. Then T is A-generated and lies in V, since V is closed under $\mathbb{P}_{\text{fin}}$ and $\mathbb{S}$. Furthermore, the following diagram commutes:



Thus T is more general than both S_1 and S_2 as an A-generated member of V. Furthermore, T is the smallest such member of V. We could iterate this process, but, as our discussion of N shows, we will never find an element of V that is more general than all other members of V. A limiting process is needed.

PROJECTIVE LIMITS

Directed set

A partially ordered set $(I, \leqslant)$ is a *directed set* if every pair of elements of I have an upper bound. [Notice that a directed set is not necessarily a join semilattice, because some pairs of elements might not have *least* upper bounds.]

Topological semigroup

A *topological semigroup* is a semigroup equipped with a topology such that the multiplication operation is a continuous mapping. Any semigroup can be equipped with the discrete topology and thus becomes a topological semigroup. Notice that finite semigroups are compact. For any alphabet A, an *A-generated topological semigroup* is a pair (S, φ), where S is a topological semigroup and $\varphi : A \to S$ is a map such that $\operatorname{im}\varphi$ generates a dense subsemigroup of S. We will often denote such an A-generated topological semigroup by the map $\varphi : A \to S$. A homomorphism between

A-generated topological semigroup

Homomorphisms between A-generated topological semigroups

A-generated topological semigroups $\varphi_1 : A \to S_1$ and $\varphi_2 : A \to S_2$ is a continuous homomorphism $\psi : S_1 \to S_2$ such that $\varphi_1\psi = \varphi_2$.

Projective system

A *projective system* is a collection of A-generated topological semigroups $\{\varphi_i : A \to S_i : i \in I\}$, where I is a directed set, such that for all $i, j \in I$ with $i \geqslant j$ there is a *connecting homomorphism* $\lambda_{i,j}$ from $\varphi_i : A \to S_i$ to $\varphi_j : A \to S_j$ satisfying the following properties: for each $i \in I$, the homomorphism $\lambda_{i,i}$ is the identity map; for all $i, j, k \in I$ with $i \geqslant j \geqslant k$, we have $\lambda_{i,j}\lambda_{j,k} = \lambda_{i,k}$.

Projective limit

The *projective limit* of this projective system is an A-generated topological semigroup $\Phi : A \to S$ equipped with homomorphisms Φ_i from $\Phi : A \to S$ to $\varphi_i : A \to S_i$. such that the following properties hold:

1) For all $i, j \in I$ with $i \geqslant j$, we have $\Phi_i\lambda_{i,j} = \Phi_j$.
2) If there is another A-generated topological semigroup $\Psi : A \to T$ and homomorphisms Ψ_i from $\Psi : A \to T$ to $\varphi_i : A \to S$ such that for all $i, j \in I$ with $i \geqslant j$, we have $\Psi_i\lambda_{i,j} = \Psi_j$, then there exists a homomorphism Θ from $\Psi : A \to T$ to $\Phi : A \to S$ such that $\Theta\Phi_i = \Psi_i$. That is, the diagram in Figure 8.1 commutes.

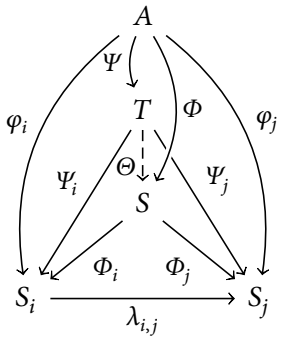


FIGURE 8.1
Property 2) of the projective limit of $\{\varphi_i : A \to S_i : i \in I\}$ with connecting homomorphisms $\lambda_{i,j}$.

Let us first show that the projective limit is unique (up to isomorphism); we will then show that it exists. Suppose $\Phi : A \to S$ and $\Phi' : A \to S'$ are both projective limits of the projective system $\{\varphi_i : A \to S_i : i \in I\}$. By property 2) above, there are homomorphisms Θ from $\Phi : A \to S$ to $\Phi' : A \to S'$ and Θ' from $\Phi' : A \to S'$ to $\Phi : A \to S$. Thus we have $\Phi\Theta\Theta' = \Phi$ and $\Phi'\Theta'\Theta = \Phi'$; hence $\Theta\Theta'|_{A\Phi} = \mathrm{id}_{A\Phi}$ and $\Theta'\Theta|_{A\Phi'} = \mathrm{id}_{A\Phi'}$. Hence $\Theta\Theta'$ restricted to the subsemigroup generated by $A\Phi$ is the identity map; since this subsemigroup is dense in S and Θ and Θ' are continuous, we have $\Theta\Theta' = \mathrm{id}_S$. Similarly $\Theta'\Theta = \mathrm{id}_{S'}$. So Θ and Θ' are mutually inverse isomorphisms between $\Phi : A \to S$ and $\Phi' : A \to S'$.

Uniqueness of the projective limit

In order to construct the projective limit, we proceed as follows. Let

Construction of the projective limit

$$S = \Big\{ s \in \prod_{i \in I} S_i : (\forall i, j \in I)\big(i \geqslant j \Rightarrow ((i)s)\lambda_{i,j} = (j)s\big) \Big\}.$$

Notice that

$$\begin{aligned}
& s, t \in S \\
\Rightarrow\ & (\forall i, j \in I)\big(i \geqslant j \Rightarrow ((i)s)\lambda_{i,j} = (j)s \wedge ((i)t)\lambda_{i,j} = (j)t\big) \\
\Rightarrow\ & (\forall i, j \in I)\big(i \geqslant j \Rightarrow ((i)s)\lambda_{i,j}((i)t)\lambda_{i,j} = (j)s(j)t\big) \\
\Rightarrow\ & (\forall i, j \in I)\big(i \geqslant j \Rightarrow ((i)(st))\lambda_{i,j} = (j)(st)\big) \\
\Rightarrow\ & st \in S;
\end{aligned}$$

thus S is a subsemigroup of $\prod_{i \in I} S_i$. Furthermore, S is equipped with the induced topology from the product topology on $\prod_{i \in I} S_i$. Let $\Phi : A \to S$ be defined by $(i)(a\Phi) = a\varphi_i$. For each $i \in I$, let Φ_i be the projection homomorphism from S to S_i.

PROPOSITION 8.11. *$\Phi : A \to S$ is an A-generated topological semigroup and satisfies the properties 1) and 2) above. Hence $\Phi : A \to S$ is a projective limit.*

Proof of 8.11. We first have to show that $A\Phi$ generates a dense subsemigroup of S. Since the topology of S is induced by the product topology on $\prod_{i\in I} S_i$, we can work with the product topology instead. Let $s \in S$. Let K be a neighbourhood of s. Assume without loss that K is an open set (in the product topology). Thus $K = \prod_{i\in I} K_i$, where each $K_i \subseteq S_i$ is open and $K_i = S_i$ for all but finitely many $i \in I$. Let $i_j \in I$ (where $j = 1, \ldots, n$) be the indices for which $K_{i_j} \neq S_{i_j}$.

Let h be an upper bound for $\{ i_j : j = 1, \ldots, n \}$; such an h exists because I is a directed set. Let $L = \bigcap_{j=1}^{n} K_{i_j} \lambda_{h,i_j}^{-1} \subseteq S_h$. Notice that $(h)s \in ((i_j)s)\lambda_{h,i_j}^{-1}$ for all $j = 1, \ldots, n$, so L contains $(h)s$ and is thus nonempty. Furthermore, L is an intersection of open sets because each $\lambda_{i,j}$ is continuous and each K_{i_j} is open; hence L is itself open. Since $A\varphi_h$ generates a dense subset of S_h, the set there is a word $w \in A^+$ such that $w\varphi_h \in L$. Let $t = w\Phi$. Thus $(i)t = w\varphi_i$ for all $i \in I$. For $j = 1, \ldots, n$, we have

$$(i_j)t = w\varphi_{i_j} = w\varphi_k \lambda_{k,i_j} \in L\lambda_{k,i_j} \subseteq K_{i_j}.$$

Hence $w\Phi = t \in K$. Therefore $A\Phi$ generates a dense subset of S.

Since S consists of elements $s \in \prod_{i\in I} S_i$ with $((i)s)\lambda_{i,j} = (j)s$, it is immediate that $\Phi_i \lambda_{i,j} = \Phi_j$; hence $\Phi : A \to S$ satisfies property 1).

Now let $\Psi : A \to T$ be an A-generated topological semigroup as described in property 2). Define $\Theta : T \to S$ by $(i)(t\Theta) = t\Psi_i$. (Note that $t\Theta \in S$ since $\Psi_i \lambda_{i,j} = \Psi_j$.) Then this map is a continuous homomorphism since each Ψ_i is continuous. Finally, $\Theta\Phi_i = \Theta\pi_i = \Psi_i$. So $\Phi : A \to S$ satisfies property 2). 8.11

Notice that if all of the S_i are compact, so is $\prod_{i\in I} S_i$ by Tychonoff's theorem. Furthermore, since each $\lambda_{i,j}$ is continuous and the condition $((i)s)\lambda_{i,j} = (j)s$ involves only two components of the product, S is closed in the $\prod_{i\in I} S_i$. Hence if all the S_i are compact, S is also compact.

Profinite semigroup

A *profinite semigroup* is a projective limit of a projective system of finite semigroups for some suitable choice of generators. Notice that any finite semigroup is [isomorphic to] a profinite semigroup. To see this, let S be a finite semigroup, and take $I = \{1\}$ and $S_1 = S$. It is easy to see that the projective limit of this projective system is isomorphic to S.

PRO-V SEMIGROUPS

Pro-V semigroup

Let V be an $\mathcal{S}$-pseudovariety. A profinite semigroup S is

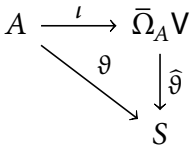
pro-V if it is a projective limit of a projective system containing only semigroups from V.

Let us return of the problem of finding a free object for V. For a generating set A, the idea is to take the projective limit of the projective system containing every A-generated semigroup in V. Strictly speaking, we take one semigroup from every isomorphism class in V, and let the connecting homomorphisms be the unique homomorphisms that respect the generating set A. The projective limit of this system is denoted $\bar{\Omega}_A\mathsf{V}$. If the A-generated semigroups in V are $\{\varphi_i : A \to S_i : i \in I\}$, then there is a natural map $\iota : A \to \bar{\Omega}_A\mathsf{V}$ given by $(i)(a\iota) = a\varphi_i$. (This is the map Φ in the discussion of the projective limit above.) Denote by $\Omega_A\mathsf{V}$ the [dense] subsemigroup of $\bar{\Omega}_A\mathsf{V}$ generated by $A\iota$.

The following result essentially says that the profinite semigroup $\bar{\Omega}_A\mathsf{V}$ is a free object for A-generated pro-V semigroups:

$\bar{\Omega}_A\mathsf{V}$ is a free object for V

PROPOSITION 8.12. *For any pro-*V *semigroup* S *and map* $\vartheta : A \to S$, *there is a unique continuous homomorphism* $\widehat{\vartheta} : \bar{\Omega}_A\mathsf{V} \to S$ *such that* $\iota\widehat{\vartheta} = \vartheta$; *that is, such that the following diagram commutes:*



Proof of 8.12. Since pro-V semigroups are subdirect products of members of V, it is sufficient to consider the case when S lies in V. Without loss of generality, assume S is generated by $A\vartheta$. Then S is [isomorphic to] an A-generated semigroup in V; that is, S is [isomorphic to] one of the A-generated semigroups $\varphi_j : A \to S_j$ in the projective system whose projective limit is $\bar{\Omega}_A\mathsf{V}$.

Let $\widehat{\vartheta}$ be the projection $\pi_j : \bar{\Omega}_A\mathsf{V} \to S_j \simeq S$. Finally, we have to show that $\widehat{\vartheta}$ is the *unique* continuous homomorphism with this property. Let $\psi : \bar{\Omega}_A\mathsf{V} \to S$ be some continuous homomorphism with $\iota\psi = \vartheta$. Since $\iota\widehat{\vartheta} = \vartheta$, we see that $\psi|_{A\iota} = \widehat{\vartheta}|_{A\iota}$ and hence, since $A\iota$ generates $\Omega_A\mathsf{V}$, we have $\psi|_{\Omega_A\mathsf{V}} = \widehat{\vartheta}|_{\Omega_A\mathsf{V}}$. Since $\Omega_A\mathsf{V}$ is dense in $\bar{\Omega}_A\mathsf{V}$ and ψ is continuous, we have $\psi = \widehat{\vartheta}$. 8.12

In light of Proposition 8.12, for any $\mathcal{S}$-pseudovariety V, we call $\bar{\Omega}_A\mathsf{V}$ the *free pro-*V *semigroup on* A.

Free pro-V semigroup

PROPOSITION 8.13. *Let* V *be an* $\mathcal{S}$*-pseudovariety that is not the trivial* $\mathcal{S}$*-pseudovariety* $\mathsf{1}$. *Then the map* $\iota : A \to \bar{\Omega}_A\mathsf{V}$ *is injective.*

Proof of 8.13. Since $\mathsf{V} \neq \mathsf{1}$, there are arbitrarily large semigroups in V. Hence for any $a, b \in A$ with $a \neq b$, there is some $\varphi_i : A \to S_i$ such that $a\varphi_i \neq b\varphi_i$. Therefore, $(i)(a\iota) = a\varphi_i \neq b\varphi_i = (i)(b\iota)$, and so $a\iota \neq b\iota$. Thus ι is injective. 8.13

Proposition 8.13 means that, when we consider any non-trivial $\mathcal{S}$-pseudovariety V, we can identify A with the subset $A\iota$ of $\overline{\Omega}_A\mathsf{V}$. From now on, assume that V is a non-trivial $\mathcal{S}$-pseudovariety.

LEMMA 8.14. *Let S be a pro-V semigroup and let $K \subseteq S$. Then the following conditions are equivalent:*

a) *there exists a continuous homomorphism $\varphi : S \to F$ such that $F \in \mathsf{V}$ and $K = K\varphi\varphi^{-1}$;*
b) *K is a clopen subset of S.*

Proof of 8.14. Suppose that condition a) holds. Then since F is finite and has the discrete topology, $K\varphi$ is clopen in F. Since φ is a continuous homomorphism, K is clopen since it is the pre-image under φ of the $K\varphi$. Thus condition b) holds.

Now suppose that condition b) holds and K is a clopen subset of S. Now, S be a subdirect product of semigroups in V. That is, S is a subsemigroup of $\prod_{i\in I} T_i$ for some $T_i \in \mathsf{V}$. Then $K = S \cap (K_1 \cup \ldots \cup K_n)$, where each K_ℓ is a product of the form $\prod_{i\in I} X_{\ell,i}$ with $X_{\ell,i} \subseteq S_i$ and $X_{\ell,i} = S_i$ for all but finitely many indices. Let

$$J = \{\, i \in I : (\exists \ell \in \{1, \ldots, n\})(T_{\ell,i} \neq S_i) \,\};$$

notice that J is finite. Let $\varphi : S \to \prod_{i\in J} S_i$ be the natural projection. Then φ is continuous, $\prod_{i\in J} S_i$ is finite, and $K = K\varphi\varphi^{-1}$. Thus condition a) holds. 8.14

PROPOSITION 8.15. *Let S be pro-V and let T be profinite. Let $\varphi : S \to T$ be a continuous homomorphism. Then* $\operatorname{im}\varphi$ *is pro-V and belongs to V if it is finite.*

Proof of 8.15. Since T is a subdirect product of finite semigroups, it is sufficient to consider the case where T is finite and φ is surjective and show that $T \in \mathsf{V}$.

For each $t \in T$, let $K_t = t\varphi^{-1}$. Then every K_t is a pre-image of a clopen set under the continuous homomorphism φ and so is clopen. By Lemma 8.14, there is, for each $t \in T$, a continuous homomorphism $\psi_t : S \to F_t$ with $F_t \in \mathsf{V}$ such that $K_t\psi_t\psi_t^{-1} = K_t$. Let $F = \prod_{t\in T} F_t$; notice that $F \in \mathsf{V}$ since T is finite. Let $\psi : S \to F$ be defined by $(t)(x\psi) = x\psi_t$. Then $\ker\psi \subseteq \ker\varphi$. Hence there is a homomorphism $\vartheta : \operatorname{im}\psi \to T$ given by $(x\psi)\vartheta = x\varphi$. Since φ is surjective, ϑ is a surjective homomorphism from the subsemigroup $\operatorname{im}\psi$ of F to the semigroup T. Hence $T \preccurlyeq F$ and so $T \in \mathsf{V}$. 8.15

Propositions 8.12, 8.13, and 8.15 together show that $\overline{\Omega}_A\mathsf{V}$ is a very good analogue for pseudovarieties of free algebras for varieties: maps from A can be extended to homomorphisms from $\overline{\Omega}_A\mathsf{V}$, the 'basis' A (usually) embeds in $\overline{\Omega}_A\mathsf{V}$, and, finally, the only finite semigroups that are homomorphic images of $\overline{\Omega}_A\mathsf{V}$ are the semigroups in V.

PSEUDOIDENTITIES

Earlier in this chapter, we saw how varieties of $\mathcal{T}$-algebras, and in particular varieties of semigroups, can be defined using laws. Recall that a law in a variety V of $\mathcal{T}$-algebras is a pair $u, v \in F_{\mathcal{T}}(A)$, usually written as a formal equality $u = v$, and that a $\mathcal{T}$-algebra S satisfies this law if $u\widehat{\varphi} = v\widehat{\varphi}$ for all homomorphisms $\widehat{\varphi} : F_{\mathcal{T}}(A) \to S$ extending maps $\varphi : A \to S$. Now that we have free objects for $\mathcal{S}$-pseudovarieties available, we can the study the analogue of laws for finite semigroups.

Pseudoidentities

Let V be an $\mathcal{S}$-pseudovariety. A V-*pseudoidentity* is a pair $u, v \in \bar{\Omega}_A\mathsf{V}$, usually written as a formal equality $u = v$. A pro-V semigroup S *satisfies* this pseudoidentity if, for every continuous homomorphism $\vartheta : \bar{\Omega}_A\mathsf{V} \to S$ we have $u\vartheta = v\vartheta$.

Now let V be an $\mathcal{M}$-pseudovariety. Then $\bar{\Omega}_A\mathsf{V}$ also exists, with the corresponding properties, and is a monoid. So we also have V-pseudoidentities $u = v$ in this case, where $u, v \in \bar{\Omega}_A\mathsf{V}$, and here u or v may be the identity of $\bar{\Omega}_A\mathsf{V}$. In this context, a pro-V monoid M satisfies this pseudoidentity if, for every continuous *monoid* homomorphism $\vartheta : \bar{\Omega}_A\mathsf{V} \to M$ we have $u\vartheta = v\vartheta$.

LEMMA 8.16. *Let* V *and* W *be* $\mathcal{S}$*-pseudovarieties* (*respectively,* $\mathcal{M}$*-pseudovarieties*) *with* $\mathsf{W} \subseteq \mathsf{V}$ *and let* $\pi : \bar{\Omega}_A\mathsf{V} \to \bar{\Omega}_A\mathsf{W}$ *be the natural projection homomorphism* (*respectively, monoid homomorphism*). *Then for any* $u, v \in \bar{\Omega}_A\mathsf{V}$, *every semigroup in* W *satisfies* $u = v$ *if and only if* $u\pi = v\pi$. 8.16

Let Σ be a set of V-pseudoidentities. Let $[\![\Sigma]\!]_{\mathsf{V}}$ denote the class of all $S \in \mathsf{V}$ that satisfy all the V-pseudoidentities in Σ.

REITERMAN'S THEOREM 8.17. *Let* $\mathcal{W}$ *be a subclass of a* $\mathcal{S}$*-pseudovariety* (*respectively,* $\mathcal{M}$*-pseudovariety*) V. *Then* $\mathcal{W}$ *is an* $\mathcal{S}$*-pseudovariety* (*respectively,* $\mathcal{M}$*-pseudovariety*) *if and only if* $\mathcal{W} = [\![\Sigma]\!]_{\mathsf{V}}$ *for some set* Σ *of* V*-pseudoidentities.*

Reiterman's theorem

Proof of 8.17. We prove the result for $\mathcal{S}$-pseudovarieties; the same reasoning works for $\mathcal{M}$-pseudovarieties with the standard modifications.

Part 1. Suppose $\mathcal{W} = [\![\Sigma]\!]_{\mathsf{V}}$. By reasoning parallel to the proof of Theorem 8.2, we see that $\mathcal{W}$ is closed under $\mathbb{H}$, $\mathbb{S}$, and $\mathbb{P}_{\text{fin}}$ and is thus an $\mathcal{S}$-pseudovariety.

Part 2. Suppose $\mathcal{W}$ is an $\mathcal{S}$-pseudovariety. Fix a countably infinite alphabet A. Let Σ be the set of all V-pseudoidentities $u = v$ satisfied by all semigroups in $\mathcal{W}$, where $u, v \in \bar{\Omega}_B\mathsf{V}$ and $B \subseteq A$. Clearly $\mathcal{W} \subseteq [\![\Sigma]\!]_{\mathsf{V}}$; we aim to prove equality.

Let $\mathsf{X} = [\![\Sigma]\!]_{\mathsf{V}}$ and let $S \in \mathsf{X}$. Then since A is infinite and S is finite, there exists some $B \subseteq A$ and a surjective continuous homomorphism $\varphi : \bar{\Omega}_B\mathsf{X} \to S$. Let $\pi : \bar{\Omega}_B\mathsf{X} \to \bar{\Omega}_B\mathsf{W}$ be the natural projection.

Suppose $u, v \in \bar{\Omega}_B \mathsf{X}$ are such that $u\pi = v\pi$. Then by Lemma 8.16, every semigroup in W satisfies $u = v$. Thus $u = v$ is a V-pseudoidentity in Σ; and thus S satisfies $u = v$. In particular, $u\varphi = v\varphi$. This shows that $\ker \pi \subseteq \ker \varphi$.

Therefore the map $\psi : \bar{\Omega}_B \mathsf{W} \to S$ defined by $(x\pi)\psi = x\varphi$ is a well-defined surjective homomorphism.

For any subset K of S, the subset $K\varphi^{-1}$ of $\bar{\Omega}_B \mathsf{X}$ is closed because φ is continuous. The map π maps closed sets to closed sets because it is a projection of compact spaces. Hence $K\psi^{-1} = K\varphi^{-1}\pi$ is closed. Thus ψ is continuous.

By Proposition 8.15, $S \in \mathsf{W}$. Therefore $[\![\Sigma]\!]_{\mathsf{V}} = \mathsf{X} \subseteq \mathcal{W}$. 8.17

Bases of pseudoidentities

If V is an $\mathcal{S}$-pseudovariety (respectively $\mathcal{M}$-pseudovariety) and Σ is a set of S-pseudoidentities (respectively, M-pseudoidentities) such that $\mathsf{V} = [\![\Sigma]\!]_{\mathsf{S}}$ (respectively, $\mathsf{V} = [\![\Sigma]\!]_{\mathsf{M}}$), then Σ is called a *basis of pseudoidentities* for V. If there is a finite set of pseudoidentities Σ such that $\mathsf{V} = [\![\Sigma]\!]_{\mathsf{S}}$ (respectively, $[\![\Sigma]\!]_{\mathsf{M}}$), then V is *finitely based*.

Notation for pseudoidentities

In order to actually write down useful pseudoidentities, we introduce some new concepts and notation. Let T be a finite semigroup, $x \in T$, and $i \in \mathbb{Z}$. Consider the sequence $(x^{n!+i})_n$. This sequence is eventually constant: for all $n > \max\{|i|, |T|\}$, all terms $x^{n!+i}$ are equal. More generally, let T be a profinite semigroup. Then the sequence $(x^{n!+i})_n$ converges to a limit, which we denote $x^{\omega+i}$. In particular, this holds when T is $\bar{\Omega}_A \mathsf{S}$ and $x \in A$.

Let S be finite and let $\vartheta : \bar{\Omega}_A \mathsf{S} \to S$ be a continuous homomorphism, the powers of $x\vartheta$ are not all distinct: we have $(x\vartheta)^{m+k} = (x\vartheta)^m$ for some $m, k \in \mathbb{N}$. Let $(x\vartheta)^n$ be the identity of the cyclic group $C = \{(x\vartheta)^m, \ldots, (x\vartheta)^{m+k-1}\}$. Since $(x\vartheta)^n = (x\vartheta)^{m!} = x^{m!}\vartheta$ for all $m \geqslant n$, we have $(x^\omega)\vartheta = (x\vartheta)^n$. That is, $(x^\omega)\vartheta$ is the unique idempotent power of $x\vartheta$. Furthermore, $x^{\omega-1}\vartheta$ is the inverse of $x^{\omega+1}\vartheta$ in C.

We can interpret this notation in S by define new operations $^{\omega+i}$ on finite semigroups. For any finite semigroup S, the operation $^\omega : S \to S$ takes any element y to its unique idempotent power y^ω. For any $k \in \mathbb{N}$, the operation $^{\omega+k} : S \to S$ takes any element y to $y^\omega y^k$, and $^{\omega-k} : S \to S$ takes y to the inverse of $y^{\omega+k}$ in the [finite] cyclic subgroup $\{y^\omega, y^{\omega+1}, \ldots\}$.

We can now give explicit examples of pseudoidentities defining particular pseudovarieties of finite semigroups and monoids. See Tables 8.4 and 8.5 for a summary.

EXAMPLE 8.18. a) The $\mathcal{S}$-pseudovariety of all finite aperiodic semigroups A is defined by the pseudoidentity $x^{\omega+1} = x^\omega$.

b) The $\mathcal{S}$-pseudovariety of finite nilpotent semigroups N is defined by the pseudoidentities $yx^\omega = x^\omega$ and $x^\omega y = x^\omega$. These pseudoidentities essentially say that $x^\omega\vartheta$ is a zero, and so we abbreviate them by $x^\omega = 0$.

c) The $\mathcal{S}$-pseudovariety of finite groups G is defined by the S-pseudoidentities $yx^\omega = y$ and $x^\omega y = y$. Since these S-pseudoidentities

Pseudovariety	Symbol	Pseudoidentities	See also
Semigroups	S	—	
Trivial semigroup	1	$x = y$	
Null semigroups	Z	$xy = zt$	
Nilpotent semigroups	N	$x^\omega = 0$	Exa. 8.18(b)
Left zero semigroups	LZ	$xy = x$	
Right zero semigroups	RZ	$xy = y$	
Rectangular bands	RB	$xyx = x$	Exer. 8.4
Comp. simple sgrps	CS	$x^{\omega+1} = x$	Exer. 8.8
Comp. regular sgrps	CR	$(xy)^\omega x = x$	Exer. 8.9
Left simple sgrps	LS	$xy^\omega = x$	Exer. 8.10
Right simple sgrps	RS	$y^\omega x = x$	Exer. 8.10
Left-trivial sgrps	K	$x^\omega y = x^\omega$	pp. 193–194
Right-trivial sgrps	D	$yx^\omega = x^\omega$	pp. 193–194

TABLE 8.4
$\mathcal{S}$-pseudovarieties of semigroups. The pseudoidentity $x(yz) = (xy)z$ is implicitly assumed in every case.

Pseudovariety	Symbol	Pseudoidentities	See also
Monoids	M	—	
Trivial monoid	1	$x = 1$	
Commutative monoids	Com	$xy = yx$	
Semilattices with ident.	Sl	$\begin{cases} x^2 = x, \\ xy = yx \end{cases}$	
Aperiodic monoids	A	$x^{\omega+1} = x^\omega$	Exa. 8.18(a)
$\mathcal{L}$-trivial monoids	L	$y(xy)^\omega = (xy)^\omega$	Pr. 8.20(a)
$\mathcal{R}$-trivial monoids	R	$(xy)^\omega x = (xy)^\omega$	Pr. 8.20(b)
$\mathcal{J}$-trivial monoids	J	$\begin{cases} (xy)^\omega x = (xy)^\omega, \\ y(xy)^\omega = (xy)^\omega \end{cases}$	Pr. 8.20(c)
Comp. regular monoids	CR	$(xy)^\omega x = x$	Exer. 8.9
Groups	G	$x^\omega = 1$	Exa. 8.18(c)
Abelian groups	Ab	$\begin{cases} xy = yx, \\ x^\omega = 1 \end{cases}$	

TABLE 8.5
$\mathcal{M}$-pseudovarieties of monoids. The pseudoidentities $x(yz) = (xy)z$, $x1 = 1$, and $1x = x$ are implicitly assumed in every case.

essentially say that $x^\omega\vartheta$ (where $\vartheta : \bar{\Omega}_A\mathsf{S} \to S$ is a homomorphism) is an identity, we abbreviate them by $x^\omega = 1$. If we consider the $\mathcal{M}$-pseudovariety of finite groups instead, then $x^\omega = 1$ is a genuine M-pseudoidentity.

We now have two different ways to define pseudovarieties: we can specify a set of $\mathcal{S}$- or $\mathcal{M}$-pseudoidentities and consider the $\mathcal{S}$-or $\mathcal{M}$-pseudovarieties of semigroups or monoids they define, or we can specify a set of finite semigroups or monoids and consider the $\mathcal{S}$- or $\mathcal{M}$-pseudovariety they generate. These ways of defining pseudovarieties interact with the lattices of pseudovarieties in different but complementary ways.

If Σ and T are sets of S-pseudoidentities, then

$$[\![\Sigma]\!]_{\mathsf{S}} \sqcap [\![T]\!]_{\mathsf{S}} = [\![\Sigma \cup T]\!]_{\mathsf{S}}, \tag{8.8}$$

and similarly for M-pseudoidentities. For example, the $\mathcal{M}$-pseudovariety of finite Abelian groups is

$$\mathsf{Ab} = [\![xy = yx]\!]_{\mathsf{M}} \sqcap [\![x^\omega = 1]\!]_{\mathsf{M}} = [\![xy = yx, x^\omega = 1]\!]_{\mathsf{M}}.$$

On the other hand, for any classes $\mathcal{X}$ and $\mathcal{Y}$ of finite semigroups,

$$\mathsf{V}_{\mathcal{S}}(\mathcal{X}) \sqcup \mathsf{V}_{\mathcal{S}}(\mathcal{Y}) = \mathsf{V}_{\mathcal{S}}(\mathcal{X} \cup \mathcal{Y}), \tag{8.9}$$

and similarly for finite monoids.

Furthermore, the operators ${}_{\mathrm{Sg}}$ and ${}_{\mathrm{Mon}}$ interact with pseudoidentities in a pleasant way. Let Σ be a set of $\mathcal{M}$-pseudoidentities. Let Σ_{Sg} be the set of $\mathcal{S}$-pseudoidentities that can be obtained from Σ as follows:

1) by substituting 1 for some of the variables;
2) replacing pseudoidentities of the form $u = 1$, where u is not the identity, by $ux = x$ and $xu = x$, where x is a new symbol not in u;
3) deleting the pseudoidentity $1 = 1$ if it is present.

PROPOSITION 8.19. a) *Let Σ be a set of $\mathcal{S}$-pseudoidentities. Then*

$$[\![\Sigma]\!]_{\mathsf{M}} = ([\![\Sigma]\!]_{\mathsf{S}})_{\mathrm{Mon}}.$$

b) *Let Σ be a set of $\mathcal{M}$-pseudoidentities. Then*

$$[\![\Sigma_{\mathrm{Sg}}]\!]_{\mathsf{S}} = ([\![\Sigma]\!]_{\mathsf{M}})_{\mathrm{Sg}}.$$

Proof of 8.19. a) Let S be a finite monoid. Then

$$\begin{aligned} & S \in [\![\Sigma]\!]_{\mathsf{M}} \\ \Leftrightarrow\ & S \text{ is a monoid that satisfies all the} \\ & \qquad \mathcal{S}\text{-pseudoidentities in } \Sigma \\ \Leftrightarrow\ & S \text{ is a monoid that belongs to } [\![\Sigma]\!]_{\mathsf{S}} \\ \Leftrightarrow\ & S \in ([\![\Sigma]\!]_{\mathsf{S}})_{\mathrm{Mon}}. \end{aligned}$$

b) For brevity, let $\mathsf{V} = [\![\Sigma]\!]_{\mathsf{M}}$. It is clear that $\mathsf{V} = [\![\Sigma_{\mathrm{Sg}}]\!]_{\mathsf{M}}$, and so $\mathsf{V}_{\mathrm{Sg}} \subseteq [\![\Sigma']\!]_{\mathsf{S}}$. Conversely, if S satisfies all the $\mathcal{S}$-pseudoidentities in Σ_{Sg}, then S^1 satisfies all the $\mathcal{M}$-pseudoidentities in Σ_{Sg}. Thus $S \in [\![\Sigma_{\mathrm{Sg}}]\!]_{\mathsf{S}}$ implies $S^1 \in \mathsf{V}$, which implies $S \in \mathsf{S}_{\mathrm{Sg}}$ by (8.7). Hence $[\![\Sigma']\!]_{\mathsf{S}} \subseteq \mathsf{V}_{\mathrm{Sg}}$. 8.19

Proposition 8.19 allows us to switch from $\mathcal{M}$-pseudoidentities for an $\mathcal{M}$-pseudovariety of monoids V to $\mathcal{S}$-pseudoidentities for corresponding monoidal $\mathcal{S}$-pseudovariety of semigroups V_{Sg}.

A semigroup S is *$\mathcal{H}$-trivial* (respectively, *$\mathcal{L}$-trivial*, *$\mathcal{R}$-trivial*, *$\mathcal{D}$-trivial*, *$\mathcal{J}$-trivial*) if $\mathcal{H}$ (respectively, $\mathcal{L}$, $\mathcal{R}$, $\mathcal{D}$, $\mathcal{J}$) is the identity relation id_S. A finite semigroup is $\mathcal{H}$-trivial if and only if it is aperiodic by Proposition 7.4, and is $\mathcal{D}$-trivial if and only if it is $\mathcal{J}$-trivial by Proposition 3.3. In particular, therefore, the class of $\mathcal{H}$-trivial finite monoids is the $\mathcal{M}$-pseudovariety A. Let L, R, and J be, respectively, the classes of $\mathcal{L}$-, $\mathcal{R}$-, and $\mathcal{J}$-trivial monoids.

PROPOSITION 8.20. a) *The class* L *is an $\mathcal{M}$-pseudovariety of monoids, and*

$$\mathsf{L} = [\![y(xy)^\omega = (xy)^\omega]\!]_{\mathsf{M}}.$$

b) *The class* R *is an $\mathcal{M}$-pseudovariety of monoids, and*

$$\mathsf{R} = [\![(xy)^\omega x = (xy)^\omega]\!]_{\mathsf{M}}.$$

c) *The class* J *is an $\mathcal{M}$-pseudovariety of monoids, and*

$$\begin{aligned}\mathsf{J} = \mathsf{L} \sqcap \mathsf{R} &= [\![y(xy)^\omega = (xy)^\omega, (xy)^\omega x = (xy)^\omega]\!]_{\mathsf{M}} \\ &= [\![(xy)^\omega = (yx)^\omega, x^\omega = x^{\omega+1}]\!]_{\mathsf{M}}.\end{aligned}$$

Proof of 8.20. a) Let $S \in \mathsf{L}$; that is, S is a finite $\mathcal{L}$-trivial monoid. Then S is $\mathcal{H}$-trivial and so aperiodic, and therefore $z^\omega = z^{\omega+1}$ for all $z \in S$. Let $x, y \in S$. Then $x(y(xy)^\omega) = (xy)^\omega$ and so $y(xy)^\omega \mathrel{\mathcal{L}} (xy)^\omega$ and so $y(xy)^\omega = (xy)^\omega$ since $\mathcal{L} = \mathrm{id}_S$. So every finite $\mathcal{L}$-trivial monoid satisfies the pseudoidentity $y(xy)^\omega = (xy)^\omega$.

On the other hand, let S be a finite monoid satisfying the pseudoidentity $y(xy)^\omega = (xy)^\omega$. Let $z, t \in S$ be such that $z \mathrel{\mathcal{L}} t$. Then there exist $p, q \in S$ such that $pz = t$ and $qt = z$. Then $t = (pq)t = (pq)^2 t = \dots$ and so $t = (pq)^\omega t$. Similarly $z = qt = q(pq)t = q(pq)^2 t = \dots$ and so $z = q(pq)^\omega t$. Substitute p for x and q for y in the pseudoidentity to see that $t = (pq)^\omega t = q(pq)^\omega t = z$. So $\mathcal{L} = \mathrm{id}_S$. Thus S is $\mathcal{L}$-trivial and so $S \in \mathsf{L}$.

Thus the class of finite $\mathcal{L}$-trivial monoids L is a pseudovariety, and $\mathsf{L} = [\![y(xy)^\omega = (xy)^\omega]\!]_{\mathsf{S}}$.

b) The reasoning is dual to part a).

c) A monoid is $\mathcal{D}$-trivial if and only if it is both $\mathcal{L}$-trivial and $\mathcal{R}$-trivial, and a finite monoid is $\mathcal{J}$-trivial if and only if it is $\mathcal{D}$-trivial. Thus $\mathsf{J} = \mathsf{L} \sqcap \mathsf{R}$, and $\mathsf{L} \sqcap \mathsf{R} = [\![y(xy)^\omega = (xy)^\omega, (xy)^\omega x = (xy)^\omega]\!]_{\mathsf{M}}$ by (8.8).

Suppose $S \in [\![y(xy)^\omega = (xy)^\omega, (xy)^\omega x = (xy)^\omega]\!]_{\mathsf{M}}$. Putting $x = y$ in the first pseudoidentity shows that $x(x^2)^\omega = (x^2)^\omega$. Since $x^\omega = (x^2)^\omega$, it follows that $x^\omega = x^{\omega+1}$ for all $x \in S$. Let $x, y \in S$ and let n be large enough that $(xy)^\omega = (xy)^n$ and $(yx)^\omega = (yx)^n$. Then $(yx)^\omega y = (yx)^n y = y(xy)^n = y(xy)^\omega$. So S satisfies the pseudoidentity $(xy)^\omega = (yx)^\omega$. Hence $S \in [\![(xy)^\omega = (yx)^\omega, x^\omega = x^{\omega+1}]\!]_{\mathsf{M}}$.

Now suppose $S \in [\![(xy)^\omega = (yx)^\omega, x^\omega = x^{\omega+1}]\!]_{\mathsf{M}}$. Let $x, y \in S$. Using both pseudoidentities, we see that $(xy)^\omega = (xy)^{\omega+1} = (yx)^{\omega+1} =$

$y(xy)^\omega x$. Hence $(xy)^\omega = y^2(xy)^\omega x^2 = y^3(xy)^\omega x^3 = \dots$ and so $(xy)^\omega = y^\omega(xy)^\omega x^\omega = y^{\omega+1}(xy)^\omega x^\omega = y(xy)^\omega$. Similarly, $(xy)^\omega = (xy)^\omega x$. Therefore $S \in [\![y(xy)^\omega = (xy)^\omega, (xy)^\omega x = (xy)^\omega]\!]_{\mathsf{S}}$. Thus

$$\begin{aligned}[\![y(xy)^\omega = (xy)^\omega, (xy)^\omega x = (xy)^\omega]\!]_{\mathsf{M}} \\ = [\![(xy)^\omega = (yx)^\omega, x^\omega = x^{\omega+1}]\!]_{\mathsf{M}}.\end{aligned} \qquad \boxed{8.20}$$

We now define another operator that connects $\mathcal{M}$-pseudovarieties of monoids with $\mathcal{S}$-pseudovarieties of semigroups. For any $\mathcal{M}$-pseudovariety of monoids V, let

$$\mathbb{L}\mathsf{V} = \{ S \in \mathsf{S} : (\forall e \in E(S))(eSe \in \mathsf{V}) \}.$$

Local submonoid, locally V

For any semigroup S and $e \in E(S)$, the subset eSe forms a submonoid whose identity is e, called the *local submonoid of S at e*. Thus a semigroup in $\mathbb{L}\mathsf{V}$ is said to be *locally* V.

Let Σ be a basis of M-pseudoidentities for a $\mathcal{M}$-pseudovariety of monoids V. Let z be a new symbol that does not appear in Σ. Let Σ' be the set of S-pseudovarieties obtained by substituting $z^\omega x z^\omega$ for x in every M-pseudoidentity in Σ, for every symbol x that appears in Σ, and substituting z^ω for 1 in every M-pseudoidentity in Σ. Then $\mathbb{L}\mathsf{V} = [\![\Sigma']\!]_{\mathsf{S}}$, and so $\mathbb{L}\mathsf{V}$ is an $\mathcal{S}$-pseudovariety of semigroups.

In particular, the variety of *locally trivial* semigroups is

$$\mathbb{L}\mathsf{1} = [\![z^\omega x z^\omega = z^\omega]\!]_{\mathsf{S}};$$

we will study these further in the next chapter.

SEMIDIRECT PRODUCT OF PSEUDOVARIETIES

Semidirect product of pseudovarieties

The *semidirect product* of two $\mathcal{S}$-pseudovarieties V and W, denoted $\mathsf{V} \rtimes \mathsf{W}$, is the $\mathcal{S}$-pseudovariety generated by all semidirect products $S \rtimes_\varphi T$, where $S \in \mathsf{V}$, $T \in \mathsf{W}$, and $\varphi : T \to \mathrm{End}(S)$ is an anti-homomorphism. We will not explore semidirect products of pseudovarieties in detail; we mention only the following result, which allows us to re-state the Krohn–Rhodes theorem in a more elegant form:

PROPOSITION 8.21. *The semidirect product of $\mathcal{S}$-pseudovarieties is associative.*

Proof of 8.21. [Technical, and omitted.] $\boxed{8.21}$

Notice that it is the semidirect product of *$\mathcal{S}$-pseudovarieties* that is associative. There is no natural definition for the associativity of the semidirect

product of semigroups: by the definition of semidirect products (see pages 133–134), the expression $(S \rtimes_\varphi T) \rtimes_\psi U$ only makes sense if the map ψ is an anti-homomorphism from U to $\mathrm{End}(S \rtimes \varphi T)$, whereas the expression $S \rtimes_\varphi (T \rtimes_\psi U)$ only makes sense if the map ψ is an anti-homomorphism from U to $\mathrm{End}(T)$.

The Krohn–Rhodes theorem shows that every finite semigroup is a wreath product of its subgroups and copies of the aperiodic semigroup U_3. Now, if V and W are $\mathcal{S}$-pseudovarieties and $S \in \mathsf{V}$ and $T \in \mathsf{W}$, then $S^T \in \mathsf{V}$ (since V is closed under finitary direct products); hence $S \wr T \in \mathsf{V} \rtimes \mathsf{W}$. Notice furthermore that $\mathsf{V} \subseteq \mathsf{V} \rtimes \mathsf{W}$ since every pseudovariety contains the trivial semigroup. Therefore the Krohn–Rhodes theorem can be restated in terms of $\mathcal{S}$-pseudovarieties as

$$\mathsf{S} = \bigcup_{k \in \mathbb{N} \cup \{0\}} \mathsf{G} \overbrace{\rtimes \mathsf{A} \rtimes \mathsf{G} \rtimes \cdots \rtimes \mathsf{A} \rtimes \mathsf{G}}^{\text{`}\rtimes \mathsf{A} \rtimes \mathsf{G}\text{' appears } k \text{ times}}.$$

EXERCISES

[See pages 241–248 for the solutions.]

8.1 Let S be cancellative semigroup that satisfies a law $u = v$ where $u, v \in A^+$ and u and v are not equal words. Without loss of generality, assume $|u| \leqslant |v|$. Let $w \in A^*$ be the longest common suffix of u and v. (That is, $u = u'w$ and $v = v'w$, where u' and v' do not end with the same letter.) Prove that

a) if $u = w$, then S is a group;

b) if $u \neq w$ then S is group-embeddable.

✶8.2 a) Prove, directly from the definition, that the class of finite nilpotent semigroups is a pseudovariety.

b) Prove that the class all nilpotent semigroups is not a variety.

✶8.3 Recall that a semigroup is *orthodox* if it is regular and its idempotents form a subsemigroup. Prove that the class of orthodox completely regular semigroups forms a variety and that it is defined by the laws $xx^{-1} = x^{-1}x$ and $xyy^{-1}x^{-1}xy = xy$.

8.4 Let RB be the class of rectangular bands.

a) Prove, directly from the definition, that RB is a variety.

b) Prove that RB is defined by the law $xyx = x$.

c) Prove that RB is also defined by the laws $x^2 = x$ and $xyz = xz$.

d) Give an example of a semigroup that satisfies $xyz = xz$ but is not a rectangular band.

8.5 Let $\mathcal{X}$ be the class of semigroups isomorphic to a direct product of a group and a rectangular band. Prove that $\mathcal{X}$ is a variety and is defined by the laws $xx^{-1} = x^{-1}x$ and $x^{-1}yy^{-1}x = x^{-1}x$.

8.6 Let $\mathcal{T}$ be a type, and let $\{ \mathsf{V}_i : i \in I \}$ be a collection of pseudovarieties of $\mathcal{T}$-algebras. Prove that $\bigcap_{i \in I} \mathsf{V}_i$ is a pseudovariety.

8.7 Prove that $(\mathsf{V}_{\mathrm{Mon}})_{\mathrm{Sg}} \subseteq \mathsf{V}$ for any $\mathcal{S}$-pseudovariety of semigroups V. Give an example to show that the inclusion may be strict.

✶8.8 Prove that the pseudovariety of finite completely regular semigroups CR is $[\![x^{\omega+1} = x]\!]_{\mathsf{S}}$.

✶8.9 Prove that the pseudovariety of finite completely simple semigroups CS is $[\![(xy)^{\omega}x = x]\!]_{\mathsf{S}}$.

8.10 Prove that $[\![xy^{\omega} = x]\!]_{\mathsf{S}}$ is the class of finite left simple semigroups. [Dual reasoning shows that $[\![y^{\omega}x = x]\!]_{\mathsf{S}}$ is the class of finite right simple semigroups.]

NOTES

The number of non-isomorphic nilpotent semigroups of order 8 is from Distler, 'Classification and Enumeration of Finite Semigroups', Table A.4. ◆ The section on varieties follows Howie, *Fundamentals of Semigroup Theory*, § 4.3 in outline, but in a universal algebraic context instead of the restricted context of $\{(\circ, 2), (^{-1}, 1)\}$-algebras. ◆ The exposition of pseudovarieties contains elements from Almeida, *Finite Semigroups and Universal Algebra*, §§ 3.1, 5.1, & 7.1. ◆ The discussion of free objects for pseudovarieties, profinite semigroups, pro-V semigroups, and pseudoidentities is based on Almeida, 'Profinite semigroups and applications' and Almeida, *Finite Semigroups and Universal Algebra*, ch. 3. ◆ For a proof of Proposition 8.21, see Almeida, *Finite Semigroups and Universal Algebra*, § 10.1. ◆ For further reading, Almeida, *Finite Semigroups and Universal Algebra* is the more accessible text, and Rhodes & Steinberg, *The q-theory of Finite Semigroups* the more recent and comprehensive monograph. Pin, *Varieties of Formal Languages*, ch. 2 and Eilenberg, *Automata, Languages, and Machines* (Vol. B), ch. v give rather different treatments.

Automata & finite semigroups

9

> We do not praise automata for accurately producing all the movements they were designed to perform, because the production of these movements occurs necessarily. It is the designer who is praised

— René Descartes, *Principles of Philosophy*, Part One, § 37 (trans. John Cottingham).

This chapter explores the connection between finite semigroups and rational languages. Rational languages are sets of words that are recognized by finite automata, which are mathematical models of simple computers. After discussing the necessary background on the theory of languages and automata, we will explore its connection to the theory of finite semigroups. The goal is the Eilenberg correspondence, which associates pseudovarieties of finite semigroups to certain classes of rational languages. We will then study some consequences of this correspondence.

FINITE AUTOMATA AND RATIONAL LANGUAGES

Language

Let A be an alphabet. A *language* over A is a subset of A^*. So a language over A is a set of words with letters in A. We will be interested in a particular class of languages over A called the rational languages. To motivate the definition of this class, we first introduce finite automata.

A finite automaton is a mathematical model of a computer with a very simple form of operation: it reads an input word (a sequence of symbols over an alphabet) one symbol at a time, and either *accepts* or *rejects* this input. The automaton can be in one of a finite number of internal states at any point. As it reads a symbol, it changes its state to a new one that is dependent on its current state and the symbol it reads. If can start in one of a given set of initial states, read an input word symbol-by-symbol, and end up in one of a given set of accept states, it accepts this input.

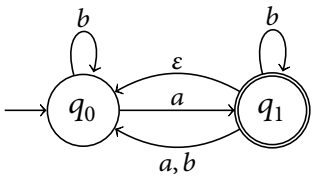
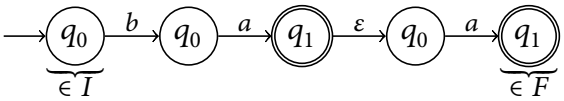


FIGURE 9.1
An example of a finite automaton.

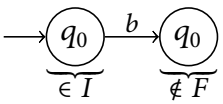
It is easier to start with an example of a finite automaton rather than a formal definition. The directed graph in Figure 9.1 represents a finite automaton. The vertices of the graph represent he states of the automaton. The state q_0 is marked with an incoming arrow 'from nowhere': this indicates that it is an initial state. The state q_1 has a double outline: this indicates that it is an accept state. (In this example, there is only one initial state and one accept state; generally there may be more than one of each, and it is possible for a state to be both an initial and an accept state.) The edges and their labels indicate how the automaton behaves: for example,

- the edge from q_0 to q_1 labelled by a says that if the automaton is in state q_0 and reads the symbol a, it can change to state q_1;
- the edge from q_1 to q_0 labelled by the empty word ε says that if the automaton is in state q_1, it can change to state q_0 without reading any symbols (that is, it can 'spontaneously' change from state q_1 to state q_0).

Notice that if the automaton is in state q_1 and reads a symbol b, it can either change to state q_0 or return to state q_1. That is, the automaton is non-deterministic: there is an element of choice in how it functions. Thus the automaton is said to accept a word $w \in \{a, b\}^*$ if there is some sequence of choices it can make so that it starts in the initial state q_0, reads w, and finishes in the accept state q_1. In terms of the graph, this is equivalent to saying that the automaton accepts w if there is a directed path in the graph starting at q_0 and ending at q_1, labelled by w. (The label on a path is the concatenation of the labels on its edges.) Hence this automaton accepts *baa*, since this word labels the path



On the other hand, it does not accept the word b, because the only path with label b starting at q_0 is the path



which does not end at q_1

Finite automaton

Formally, a *finite automaton*, or simply an *automaton*, $\mathcal{A}$ is formally a quintuple (Q, A, δ, I, F), where Q is a finite set of *states*, A is a finite

alphabet, $\delta : Q \times (A \cup \{\varepsilon\}) \to \mathbb{P}Q$ is a map called the *transition function*, $I \subseteq Q$ is a set of distinguished states called the *initial states* or *start states*, and $F \subseteq Q$ is a distinguished set of states called *accept states* or *final states*.

We think of an automaton as a directed graph with labelled edges, with vertices being the states, and, for each $q \in Q$ and $a \in A$, and for each $q' \in (q, a)\delta$, an edge labelled by a from q to q'. We can thus represent an automaton in a diagrammatic form, with the states being nodes connected by arrows. Initial states are marked with an incoming arrow 'from nowhere'. Accept states have double borders. For each $q \in Q$ and $a \in A$, there is an arrow labelled by $a \in A$ from q to each element of $(q, a)\delta$. The label on a path in such a graph is the product of the labels on the edges in that path.

For example, let $\mathcal{A}$ be automaton in Figure 9.1. Then $\mathcal{A}$ has state set $Q = \{q_0, q_1\}$. The set of initial states is $I = \{q_0\}$, the set of final states is $F = \{q_1\}$, and the transition function $\delta : Q \times (A \cup \{\varepsilon\}) \to \mathbb{P}Q$ is as given in Table 9.1.

	ε	a	b
q_0	$\varnothing$	$\{q_1\}$	$\{q_0\}$
q_1	$\{q_0\}$	$\{q_0\}$	Q

TABLE 9.1
Values of $(q, a)\delta$

Accepted word

We say that an automaton $\mathcal{A} = (Q, A, \delta, I, F)$ *accepts* a word $w \in A^*$ if there is a directed path in the diagram starting at an initial state in I and ending at an accept state in F, and labelled by w.

The idea is that the automaton is a model of a computer that can start in any state in I. While in state q, it can read a letter a from an input tape and change to any state in $(q, a)\delta$, or it can change to any state in $(q, \varepsilon)\delta$ without reading any input. The automaton accepts its input if, when it has finished reading all the input letters, it is in a state in F.

Language recognized by an automaton

Recognizable language

The set of all words accepted by an automaton $\mathcal{A}$ is denoted $L(\mathcal{A})$, and is called the *language recognized by* $\mathcal{A}$. If a language $L \subseteq A^*$ is recognized by some finite automaton, it is called a *recognizable language*.

Our description of an automaton reading input involves an element of choice. The automaton is *non-deterministic*: First, the automaton can start in any state in I. Second, the action it takes when it is in a particular state with a particular input letter to read is not fixed: the automaton can change to one of several other states on reading that letter, and may indeed change to another state without reading any input.

Deterministic automaton

An automaton where there is no such choice is called *deterministic*. More formally, an automaton $\mathcal{A} = (Q, A, \delta, I, F)$ is *deterministic* if I contains exactly one state, $\delta(q, \varepsilon) = \varnothing$ for all $q \in Q$, and $\delta(q, a)$ contains a single state for all $q \in Q$ and $a \in A$. In terms of the diagram, $\mathcal{A}$ is deterministic if there is only one state with an incoming edge 'from nowhere', no edge is labelled by ε, and for every state $q \in Q$ and $a \in A$, there is at most one edge starting at q and labelled by a. So in a deterministic automaton, there is at most one path starting at a given state and labelled by a given word. (Such a path may not exist, since there might not an edge with the required label present at some point.)

However, although deterministic automata seem to be much more restrictive than non-deterministic ones, the class of deterministic automata

actually has the same 'recognizing power' as the class of all automata, in a sense made precise by the following result:

Recognizable languages are recognized by deterministic automata

THEOREM 9.1. *Let L be a recognizable language. Then there is a deterministic automaton that recognizes L.*

Proof of 9.1. Let $\mathcal{A} = (Q, A, \delta, I, F)$ be an automaton, possibly non-deterministic, that recognizes L. For the purposes of this proof, we define the ε-closure of a set $P \subseteq Q$ to be the set

$$C_\varepsilon(P) = \{\, r \in Q : (\exists p \in P)(\text{there is a path in } \mathcal{A} \text{ from } p \text{ to } q \text{ labelled by } \varepsilon) \,\}.$$

We are going to define a new automaton $D(\mathcal{A}) = (\mathbb{P}Q, A, \eta, J, G)$. Note that the state set of $D(\mathcal{A})$ is the power set of the state set of $\mathcal{A}$. The idea is that each state of $D(\mathcal{A})$ is a set that records every possible state that $\mathcal{A}$ could be at a given time. The following definitions formalize this idea.

The set of initial states J is the singleton set $\{C_\varepsilon(I)\}$. (Note that before reading any symbol, $\mathcal{A}$ could be in any state in $C_\varepsilon(I)$.)

The transition function η has domain $Q \times (A \cup \{\varepsilon\})$ and codomain the power set of the power set of Q. The function η is defined by

$$(S, \varepsilon)\eta = \varnothing,$$
$$(S, a)\eta = \Big\{ C_\varepsilon(\{\, p \in Q : (\exists q \in S)((q, a)\delta = p) \,\}) \Big\}.$$

We emphasize that $(S, a)\eta$ contains a single element $C_\varepsilon(\ldots)$. Note that, starting in a state in S and reading a symbol a, the automaton $\mathcal{A}$ could be in any state in $(S, a)\eta$.

Finally, the set of accept states is

$$G = \{\, U \in \mathbb{P}Q : U \cap F \neq \varnothing \,\}.$$

Note that $D(\mathcal{A})$ is deterministic. We now have to prove that $L(\mathcal{A}) = L(D(\mathcal{A}))$.

Suppose $w = w_1 \cdots w_n \in L(\mathcal{A})$. Then there is a path in the graph of $\mathcal{A}$ from a state in I to a state in F. We are going to prove that there is a path in $D(\mathcal{A})$ from the (unique) initial state to an accept state with the same label. Let $q_0, \ldots, q_k$ be the states on a path in $\mathcal{A}$ labelled by w, with $q_0 \in I$ and $q_k \in F$. For $j = 0, \ldots, n-1$, let i_j the the subscript of the state immediately before the edge labelled by the symbol w_{j+1} on this path.

So there is a path in $\mathcal{A}$ from $q_0 \in I$ to q_{i_0} labelled by ε, so $q_{i_0} \in C_\varepsilon(I)$. So q_{i_0} is in the unique initial state of $D(\mathcal{A})$. Let $Q_{i_0} = C_\varepsilon(I)$; then we have $q_{i_0} \in Q_{i_0}$.

Proceed by induction. For any j, we have $q_{i_j+1} \in (q_{i_j}, w_j)\delta$, and we know that there is a path in $\mathcal{A}$ from q_{i_j+1} to $q_{i_{j+1}}$ labelled by ε. Therefore

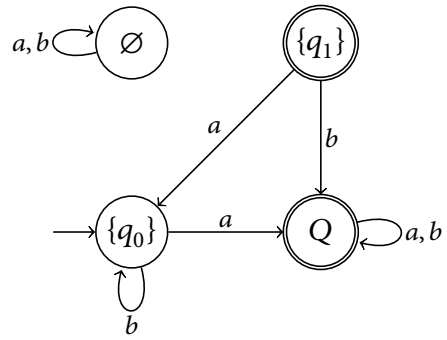


FIGURE 9.2
The deterministic automaton equivalent to the one in Figure 9.1.

we have $q_{i_{j+1}} \in C_\varepsilon(\{q_{i_j+1}\}) \subseteq (Q_{i_j}, w_j)\eta$. Let $Q_{i_{j+1}} = (Q_{i_j}, w_j)\eta$; then we have $q_{i_{j+1}} \in Q_{i_{j+1}}$.

Finally, there is a path from $q_{i_{n-1}+1}$ to $q_k \in F$ labelled by ε, so $(Q_{i_n}, w_n)\eta$ must contain q_k, and hence $(Q_{i_{n-1}}, w_n)\eta \cap F \neq \varnothing$. Let $Q_{i_n} = (Q_{i_{n-1}}, w_n)\eta$. Thus the (unique) path in $D(\mathcal{A})$ labelled by $w_1 \cdots w_n$ and starting from the (unique) initial state $Q_{i_0} \in J$ visits states $Q_{i_0}, Q_{i_0}, \ldots, Q_{i_{n-1}}, Q_{i_n}$ and ends at an accept state. So $w \in L(D(\mathcal{A}))$.

Now suppose that $w = w_1 \ldots, w_n \in L(D(\mathcal{A})$. Then there is a sequence of states $Q_0, \ldots, Q_n$, with $Q_0 \in J$ and $Q_n \in G$, and $(Q_{i-1}, w_i)\eta = Q_i$ for $i = 1, \ldots, n$. By the definition of G, there is some $q'_n \in Q_n \cap F$. Proceed by induction. For any $j = 1, \ldots, n$, by the definition of η, there exist $q_{j-1} \in Q_{j-1}$ and $q_j \in Q_j$ such that $q_j \in (q_{j-1}, w_j)\delta$ and there with a path from q_j to q'_j in $\mathcal{A}$ labelled by ε.

Finally, since $Q_0 \in J$, we have $Q_0 = C_\varepsilon(I)$ and so there is a path in $\mathcal{A}$ labelled by ε from some $q_0 \in I$ to q'_0.

Hence there is a path in $\mathcal{A}$ from $q_0 \in I$ to $q'_n \in F$ passing through $q'_0, q_1, q'_1 \ldots, q_{n-1}, q'_{n-1}, q_n$ (and other intermediate states) and labelled by $w = w_1 \cdots w_n$. So $w \in L(\mathcal{A})$.

Thus the recognizable language L is recognized by the deterministic automaton $D(\mathcal{A})$. 9.1

Applying the construction in the proof of Theorem 9.1 to the example automaton $\mathcal{A}$ above, the resulting deterministic automaton $D(\mathcal{A})$ recognizing $L(\mathcal{A})$ has set of initial states $J = \{\{q_0\}\}$, set of accept states $G = \{\{q_1\}, Q\}$, and transition function $\eta : \mathbb{P}Q \times (A \cup \{\varepsilon\}) \to \mathbb{P}(\mathbb{P}Q)$ as shown in Table 9.2. Diagrammatically, $D(\mathcal{A})$ is shown in Figure 9.2.

	ε	a	b
$\varnothing$	$\varnothing$	$\{\varnothing\}$	$\{\varnothing\}$
$\{q_0\}$	$\varnothing$	$\{Q\}$	$\{\{q_0\}\}$
$\{q_1\}$	$\varnothing$	$\{\{q_0\}\}$	Q
Q	$\varnothing$	$\{Q\}$	Q

TABLE 9.2
Values of $(S, a)\eta$

We will need to make a distinction between two classes of languages. A $*$-*language* is a subset of A^*; that is, it may include the empty word. A $+$-*language* is a subset of A^+; that is, it does not contain the empty word. Of course, every $+$-language can also be viewed as a $*$-language. But the distinction is important when we perform operations on languages, and when we develop the correspondence of classes of languages and pseudovarieties.

$*$-languages, $+$-languages

Let A be an alphabet. We are going to define some operations on the classes of languages over A. Let L and K be $*$-languages over A. Then

Boolean operations, Boolean algebra

$K \cup L$ and $K \cap L$ are, respectively, the union and intersection of K and L. The language $A^* \setminus L$ is the complement of L in A^*. Notice that the class of $*$-languages is closed under union, intersection, and complement. These are the *Boolean operations* on the class of $*$-languages. We will say that a class of $*$-languages that is closed under the Boolean operations is a *Boolean algebra*. [The notion of a Boolean algebra is more general than this, but this definition will suffice for us.]

For +-languages, we have the same union and intersection operations. However, the complement operation is different: $A^+ \setminus L$ is the complement of L in A^+, and is also a +-language. The class of +-languages is closed under the operations of union, intersection, and this new complement operation; these are the Boolean operations on the class of +-languages. Again, we say that a class of +-languages that is closed under the Boolean operations is a Boolean algebra.

Kleene star *, Kleene plus $^+$

The concatenation KL of the $*$-languages or +-languages K and L is the set of words of the form uv, where $u \in K$ and $v \in L$. The submonoid of A^* generated by K is K^*; the subsemigroup generated by K is K^+. Note that when $K = A$, this agrees with the notation for the free monoid and free semigroup. However, when $K \neq A$, the sets K^* and K^+ are in general not the free monoid and free semigroup on K. For example, if $K = \{a^2, a^3\}$, then $a^2a^3 = a^3a^2$, so K^* and K^+ are not the free monoid and free semigroup on K. The operations * and $^+$ are called the *Kleene star* and *Kleene plus*. Notice that the class of $*$-languages is closed under the operations * and $^+$, and the class of +-languages is closed under the operation $^+$.

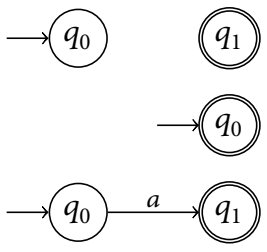
Rational/regular language

A language over $A = \{a_1, \ldots, a_n\}$ is *rational* or *regular* if it can be obtained from the languages $\varnothing$, $\{\varepsilon\}$, $\{a_1\}$, $\{a_2\}$, …, $\{a_n\}$, by applying (zero or more times) the operations of union, concatenation, and Kleene star.

Kleene's theorem

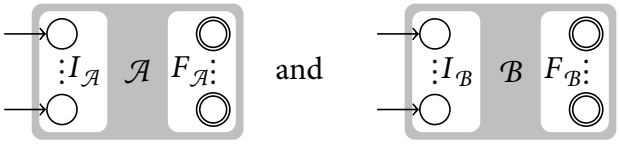
KLEENE'S THEOREM 9.2. *A language over a finite alphabet is rational if and only if it is recognizable.*

Proof of 9.2. To show that any rational language is recognizable, it suffices to show that the languages $\varnothing$, $\{\varepsilon\}$, and $\{a\}$ (for $a \in A$) are recognizable, and then to prove that the class of recognizable languages is closed under concatenation, union, and Kleene star. In our various constructions, we will simply draw automata. First, notice that

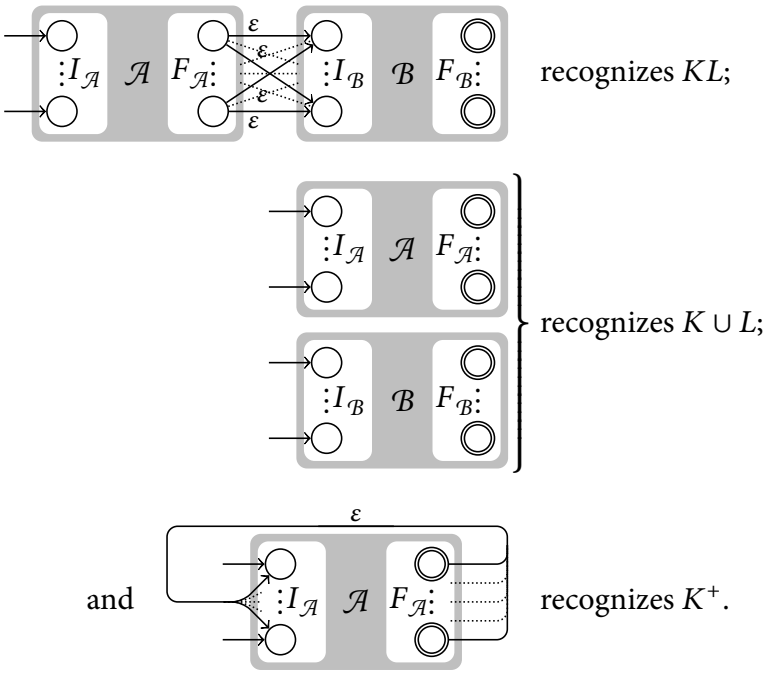


So $\varnothing$, $\{\varepsilon\}$, and $\{a\}$ (for $a \in A$) are recognizable.

Now suppose that K and L are recognizable languages. So there are automata $\mathcal{A} = (Q_{\mathcal{A}}, A, \delta_{\mathcal{A}}, I_{\mathcal{A}}, F_{\mathcal{A}})$ and $\mathcal{B} = (Q_{\mathcal{B}}, B, \delta_{\mathcal{B}}, I_{\mathcal{B}}, F_{\mathcal{B}})$ recognizing K and L respectively, which we will sketch as



respectively. Then



Thus KL, $K \cup L$, and K^+ are recognizable languages. Hence $K^* = K^+ \cup \{\varepsilon\}$ is also recognizable. This proves that every rational language is recognizable.

So let L be a recognizable language. Then by Theorem 9.1, there is a deterministic automaton $\mathcal{A} = (Q, A, \delta, I, F)$ recognizing L. Suppose that $Q = \{q_1, \ldots, q_n\}$ and $I = \{q_1\}$. For each $i, j \in \{1, \ldots, n\}$ and $k \in \{0, \ldots, n\}$, let $R[i, j; k]$ be the set of labels on paths that start at q_i, end at q_j, and visit only intermediate vertices in $\{q_1, \ldots, q_k\}$. (So $R[i, j; 0]$ is the set of labels on paths that start at q_i, end at q_j, and do not visit any intermediate vertices.) Note that

$$L = L(\mathcal{A}) = \bigcup_{\substack{1 \leqslant j \leqslant n \\ q_j \in F}} R[1, j; n]. \tag{9.1}$$

Thus we aim to prove that $R[i, j; k]$ is rational for all $i, j \in \{1, \ldots, n\}$ and $k \in \{0, \ldots, n\}$; this will suffice to prove L is rational. We proceed by induction on k.

First, consider $k = 0$. The words in $R[i, j; 0]$ label paths from q_i to q_j that visit no intermediate vertices, and so can have length at most 1 (length 0 is possible if $i = j$). So $R[i, j; 0]$ is either $\varnothing$ or a union of sets $\{\varepsilon\}$ and $\{a\}$ (for $a \in A$). This is the base of the induction.

Now let $k > 0$ and assume that $R[i, j; k-1]$ is rational for all i, j. Consider a path α from q_i to q_j that only visits intermediate vertices from $\{q_1, \ldots, q_k\}$. Now, if α does not visit q_k, then its label is in $R[i, j; k-1]$. Otherwise we can decompose α into subpaths between visits to q_k: that is, let α be the concatenation of subpaths $\alpha_0, \alpha_1, \ldots, \alpha_m$ where α_0 is the subpath from q_i up to the first visit to q_k, the α_ℓ (for $\ell = 1, \ldots, m-1$) are the subpaths between visits to q_k, and α_m is the subpath from the last visit to q_k to the state q_j. Then the label on α_0 is in $R[i, k; k-1]$, the labels on the α_ℓ (for $\ell = 1, \ldots, n-1$) are in $R[k, k; k-1]$, and the label on α_n is in $R[k, j; k-1]$. Since α was arbitrary, this shows that

$$R[i, j; k] = R[i, j; k-1] \cup R[i, k; k-1]\big(R[k, k; k-1]\big)^* R[k, j; k-1].$$

By the induction hypothesis, each of the sets $R[i, j; k-1]$ are rational. Thus $R[i, j; k]$ is rational, since it obtained from these sets using concatenation, union, and Kleene star.

Hence, by induction, all the sets $R[i, j; k]$ are rational. Therefore L is rational by (9.1). 9.2

As a consequence of Theorem 9.2, the class of rational languages is closed under complementation. Hence we may view the rational languages over A as the languages obtainable from $\{a\}$ (for $a \in A$) and $\{\varepsilon\}$ by applying the operations of union, concatenation, and Kleene star, and also intersection, complement, and Kleene plus.

Star-free/plus-free languages

A $*$-language L over A is *star-free* if it can be obtained from the languages $\{a\}$, where $a \in A$, and $\{\varepsilon\}$ using only the operations of union, intersection, complementation, and concatenation.

Left-/right-quotients of a language

For any $*$-language L over an alphabet A and word $u \in A^*$, define the languages

$$u^{-1}L = \{ w \in A^* : uw \in L \}$$
$$Lu^{-1} = \{ w \in A^* : wu \in L \};$$

the languages $u^{-1}L$ and Lu^{-1} are, respectively, the *left* and *right quotients* of L with respect to u. Similarly, for any $+$-language L over an alphabet A and word $u \in A^*$, define

$$u^{-1}L = \{ w \in A^+ : uw \in L \}$$
$$Lu^{-1} = \{ w \in A^+ : wu \in L \}.$$

Notice that the class of $+$-languages is closed under forming left and right quotients. The following result shows that the classes of *rational* $*$- and $+$-languages are also closed under forming left and right quotients:

PROPOSITION 9.3. *If L is a rational $*$-language (respectively, $+$-language), then L has only finitely many distinct left and right quotients, all of which are rational $*$-languages (respectively, $+$-language).*

Proof of 9.3. We will prove the result for $*$-language, the proof for $+$-languages is identical, with the additional observation the class of $+$-languages is closed under forming left and right quotients.

Let $\mathcal{A} = (Q, A, \delta, I, F)$ be an automaton with $L = L(\mathcal{A})$. Let $u \in A^*$.

Let $J \subseteq Q$ consist of all states $\mathcal{A}$ can reach starting at a state in I and reading u. Let ${}_J\mathcal{A} = (Q, A, \delta, J, F)$. Then $w \in L({}_J\mathcal{A})$ if and only if $uw \in L(\mathcal{A})$. That is, $u^{-1}L = u^{-1}L(\mathcal{A}) = L({}_J\mathcal{A})$. So $u^{-1}L$ is rational. Since there are only finitely many possibilities for J, there are only finitely many distinct languages $u^{-1}L$.

Similarly, let $G \subseteq Q$ consist of all states in which $\mathcal{A}$ can start and reach a state in F after reading u. Let $\mathcal{A}_G = (Q, A, \delta, I, G)$. Then $w \in L(\mathcal{A}_G)$ if and only if $wu \in L(\mathcal{A})$. That is, $Lu^{-1} = L(\mathcal{A})u^{-1} = L(\mathcal{A}_G)u^{-1}$. So Lu^{-1} is rational. Since there are only finitely many possibilities for G, there are only finitely many distinct languages Lu^{-1}. 9.3

Complete automaton

Let $\mathcal{A} = (Q, A, \delta, \{q_0\}, F)$ be a deterministic automaton. Note that for each $q \in Q$ and $a \in A$, the set $(q, a)\delta$ is either empty or contains a single element. If δ is such that $(q, a)\delta$ is never empty (that is, there is exactly one element in each $(q, a)\delta$), then the automaton $\mathcal{A}$ is *complete*. In terms of the graph, a complete deterministic automaton has exactly one edge starting at each vertex with each label in A.

Transition monoid of an automaton

Let $\mathcal{A} = (Q, A, \delta, \{q_0\}, F)$ be a complete deterministic automaton. For each $a \in A$, there is a map $\tau_a : Q \to Q$ with $q\tau_a$ given by $(q, a)\delta = \{q\tau_a\}$. Notice that $\tau_a \in \mathcal{T}_Q$ for each $a \in A$. So we have a homomorphism $\varphi : A^* \to \mathcal{T}_Q$ extending the map $a \mapsto \tau_a$. The set $\operatorname{im}\varphi$ is a submonoid of $\mathcal{T}_Q$ called the *transition monoid* of $\mathcal{A}$. For any word $w \in A^*$, the element $w\varphi$ is a transformation of Q. For any $q \in Q$, the state $q(w\varphi)$ is the state that $\mathcal{A}$ will reach if it starts in q and reads w. Let $Y = \{\sigma \in \operatorname{im}\varphi : q_0\sigma \in F\} \subseteq \mathcal{T}_Q$. Then $w\varphi \in Y$ if and only if $w \in L(\mathcal{A})$. That is, we have a monoid $\mathcal{T}_Q$ with a subset Y and a homomorphism $\varphi : A^* \to \mathcal{T}_Q$ that describes $L(\mathcal{A})$ as the inverse image of Y under φ. This motivates the following definition.

Language recognized by a homomorphism

A $*$-language L over A is *recognized by a homomorphism into a monoid M*, or more simply *recognized by a monoid M*, if there exists a monoid homomorphism $\varphi : A^* \to M$, where M is a monoid with a subset M' of M such that $L = M'\varphi^{-1}$, or, equivalently, with $L = L\varphi\varphi^{-1}$. Similarly, a $+$-language L over A is *recognized by a homomorphism into a semigroup S*, or more simply *recognized by a semigroup S*, if there exists a homomorphism $\varphi : A^+ \to S$, where S is a semigroup, with $L = L\varphi\varphi^{-1}$. Notice that if L is a $*$-language (respectively, a $+$-language) recognized by $\varphi : A^* \to M$ (respectively, $\varphi : A^+ \to S$), then $L = \varphi\varphi^{-1}$ and so $L = \bigcup_{x \in L\varphi} x\varphi^{-1}$. Each set $x\varphi^{-1}$ consists of words that map to x and so are related by $\ker\varphi$. That is, each $x\varphi^{-1}$ is a $\ker\varphi$-class, so L is a union of $\ker\varphi$-classes. Notice that

in the discussion in the previous paragraph, $\mathcal{T}_Q$ is a finite monoid. So any recognizable language is recognized by a finite monoid. Furthermore, any recognizable +-language L is recognized by a finite semigroup, since in this case the initial state of an automaton recognizing L cannot also be an accepting state, and hence Y in the discussion above does not contain id_Q, so that we can use $\varphi|_{A^+} : A^+ \to (\mathcal{T}_Q \setminus \{\mathrm{id}_Q\})$, noting that $L = Y\varphi|_{A^+}^{-1} = L\varphi|_{A^+}\varphi|_{A^+}^{-1}$.

On the other hand, suppose L is a $*$-language over A recognized by a finite monoid M. Let $\varphi : A^* \to M$ be a homomorphism recognizing L, so that by $L = L\varphi\varphi^{-1}$ Then we can construct an automaton $\mathcal{A}$ recognizing L as follows. The state set is M. The set of initial states is $\{1_M\}$, the set of accept states is $L\varphi$, and the transition function $\delta : M \times (A \cup \{\varepsilon\}) \to M$ is given by $(m, a)\delta = \{m(a\varphi)\}$. It is easy to see that $L(\mathcal{A}) = L\varphi\varphi^{-1} = L$, since the unique path starting at 1_M and labelled by $w = w_1 \cdots w_n$ (where $w_i \in A$) is

$$\longrightarrow (1_M) \xrightarrow{w_1} (w_1\varphi) \xrightarrow{w_2} ((w_1w_2)\varphi) \xrightarrow{w_3} \cdots \xrightarrow{w_n} ((w_1 \cdots w_n)\varphi)$$

This path ends in $L\varphi$ if and only if $w \in L$. Similarly, if a +-language L is recognized by a finite semigroup S, we can construct an automaton recognizing S with state set S^1. Thus we have proven the following result:

THEOREM 9.4. *A $*$-language is recognizable if and only if it is recognized by a finite monoid. A +-language is recognizable if and only if it is recognized by a finite semigroup.* 9.4

V-recognizability

Let V be an $\mathcal{M}$-pseudovariety of monoids (respectively, an $\mathcal{S}$-pseudovariety of semigroups). A $*$-language (respectively, +-language) over A is V-*recognizable* if it is recognized by some monoid (respectively, semigroup) in V. Thus Theorem 9.4 says that a $*$-language (respectively +-language) is recognizable if and only if it is M-recognizable (respectively, S-recognizable).

At this point, our goal is to describe classes of languages that are V-recognizable for a given pseudovariety V.

SYNTACTIC SEMIGROUPS AND MONOIDS

We are now going to study particular semigroups and monoids associated to languages, known as syntactic monoids and semigroups. These will be of fundamental importance in establishing a connection between pseudovarieties and classes of recognizable languages.

σ_L

For any $*$-language L over A, define a relation σ_L on A^* as follows: for $u, v \in A^*$,

$$u \,\sigma_L\, v \Leftrightarrow (\forall p, q \in A^*)(puq \in L \Leftrightarrow pvq \in L). \tag{9.2}$$

For any +-language L, define σ_L on A^+ in exactly the same way, using (9.2); note in particular that p and q still range over A^*, *not* A^+.

PROPOSITION 9.5. *Let L be a $*$-language (respectively, +-language) over A. Then:*

a) *The relation σ_L is a congruence on A^* (respectively, A^+).*
b) *The language L is a union of σ_L-classes.*
c) *If ρ is a congruence on A^* (respectively, A^+) with the property that L is a union of ρ-classes, then $\rho \subseteq \sigma_L$.*

Proof of 9.5. We prove the result for $*$-languages; the reasoning for +-languages is parallel.

a) It is immediate from the definition that σ_L is reflexive, symmetric, and transitive. So σ_L is an equivalence relation. Let $u \ \sigma_L \ v$ and let $s \in A^*$. Then $puq \in L \Leftrightarrow pvq \in L$ for all $p, q \in A^*$. In particular, this holds for all p of the form $p's$; hence $p'suq \in L \Leftrightarrow p'svq \in L$ for all $p', q \in A^*$. Hence $su \ \sigma_L \ sv$. So σ_L is left-compatible; similarly it is right-compatible and is thus a congruence.
b) Let $u \in L$ and let $v \ \sigma_L \ u$. Put $p = q = \varepsilon$ in the definition of σ_L to see that $v \in L$. Thus if any σ_L-class intersects L, it is contained in L. Therefore L is a union of σ_L-classes.
c) Let ρ be a congruence on A^* such that L is a union of ρ-classes. Then

$$\begin{aligned}
&(u, v) \in \rho \\
\Rightarrow\ &(\forall p, q \in A^*)((puq, pvq) \in \rho) && \text{[since } \rho \text{ is a congruence]} \\
\Rightarrow\ &(\forall p, q \in A^*)\big((puq, pvq) \in L \vee (puq, pvq) \notin L\big) \\
& && \text{[since } L \text{ is a union of } \rho\text{-classes]} \\
\Rightarrow\ &(\forall p, q \in A^*)(puq \in L \Leftrightarrow pvq \in L) \\
\Rightarrow\ &(u, v) \in \sigma_L;
\end{aligned}$$

thus $\rho \subseteq \sigma_L$. 9.5

Syntactic congruence

For any language L over an alphabet A, the congruence σ_L is called the *syntactic congruence* of L.

Syntactic monoid

For any $*$-language L, the factor monoid A^*/σ_L is called the *syntactic monoid* of L and is denoted $\mathrm{Syn}M\, L$, and the natural monoid homomorphism $\sigma_L^\natural : A^* \to A^*/\sigma_L = \mathrm{Syn}M\, L$ is the *syntactic monoid homomorphism* of L.

Syntactic semigroup

For any +-language L, the factor semigroup A^+/σ_L is called the *syntactic semigroup* of L and is denoted $\mathrm{Syn}S\, L$, and the natural homomorphism $\sigma_L^\natural : A^+ \to A^+/\sigma_L = \mathrm{Syn}S\, L$ is the *syntactic homomorphism* of L.

The importance of syntactic monoids and semigroups is the following result. Essentially, it says that the syntactic monoid of a $*$-language is the smallest monoid that recognizes that language, and similarly for +-languages and semigroups:

PROPOSITION 9.6. a) *Let L be a $*$-language. Then L is recognized by a monoid M if and only if* $\mathrm{SynM}\, L \preccurlyeq M$.

b) *Let L be a $+$-language. Then L is recognized by a semigroup S if and only if* $\mathrm{SynS}\, L \preccurlyeq S$.

Proof of 9.6. We prove only part a); the proof for part b) follows by replacing 'A^*' by 'A^+', 'submonoid' by 'subsemigroup', 'SynM' by 'SynS', and 'monoid homomorphism' by 'homomorphism' throughout.

Let $\varphi : A^* \to M$ recognize L. So $L = L\varphi\varphi^{-1}$. Then $\ker\varphi$ is a congruence on A^* and L is a union of $\ker\varphi$-classes. Hence $\ker\varphi \subseteq \sigma_L$. Define a map $\psi : \mathrm{im}\,\varphi \to \mathrm{SynM}\, L$ by $(u\varphi)\psi = [u]_{\sigma_L}$; this map is a well-defined monoid homomorphism since $\ker\varphi \subseteq \sigma_L$. It is clearly surjective. Since $\mathrm{im}\,\varphi$ is an $\mathcal{M}$-submonoid of M and $\psi : \mathrm{im}\,\varphi \to \mathrm{SynM}\, L$ is a surjective homomorphism, $\mathrm{SynM}\, L \preccurlyeq M$.

For the other direction, we first prove that $\mathrm{SynM}\, L$ recognizes L. Since L is a union of σ_L-classes, it follows that $L = \bigcup_{u\in L}[u]_{\sigma_L} = \bigcup_{x\in L} x(\sigma_L^\natural)^{-1} = L\sigma_L^\natural(\sigma_L^\natural)^{-1}$. Thus $\sigma_L^\natural : A^* \to \mathrm{SynM}\, L$ recognizes L.

Let $\mathrm{SynM}\, L \preccurlyeq M$. So there is an $\mathcal{M}$-submonoid N of M and a surjective $\mathcal{M}$-homomorphism $\psi : N \to \mathrm{SynM}\, L$. For each $a \in A$, define a map $\varphi : A \to N$ by choosing $a\varphi$ such that $(a\varphi)\psi = a\sigma_L^\natural$. Since A^* is free on A, there is a unique extension of φ to a homomorphism $\widehat{\varphi} : A^* \to N$; notice that $(u\widehat{\varphi})\psi = u\sigma_L^\natural$ for all $u \in A^*$ since ψ and $\sigma_L^\natural$ are monoid homomorphisms. Let $N' = L\sigma_L^\natural\psi^{-1}$. Then, viewing $\widehat{\varphi}$ as a homomorphism from A^* to M, we have

$$L = L\sigma_L^\natural(\sigma_L^\natural)^{-1} = L\sigma_L^\natural\psi^{-1}\widehat{\varphi}^{-1} = N'\widehat{\varphi}^{-1},$$

and so M recognizes L. 9.6

Proposition 9.6 is actually the original motivation behind the concept of division.

Properties of syntactic monoids

PROPOSITION 9.7. *Let A and B be alphabets. For all $*$-languages L and K over A, for all $a \in A$, and for all monoid homomorphisms $\varphi : B^* \to A^*$:*

a) $\mathrm{SynM}\, L = \mathrm{SynM}(A^* \smallsetminus L)$;

b) $\mathrm{SynM}(L \cap K) \preccurlyeq (\mathrm{SynM}\, L) \times (\mathrm{SynM}\, K)$;

c) $\mathrm{SynM}(L \cup K) \preccurlyeq (\mathrm{SynM}\, L) \times (\mathrm{SynM}\, K)$;

d) $\mathrm{SynM}(a^{-1}L) \preccurlyeq \mathrm{SynM}\, L$;

e) $\mathrm{SynM}(La^{-1}) \preccurlyeq \mathrm{SynM}\, L$;

f) $\mathrm{SynM}(L\varphi^{-1}) \preccurlyeq \mathrm{SynM}\, L$.

Proof of 9.7. a) For any $u, v \in A^*$, we have

$$\begin{aligned} u\ \sigma_L\ v &\Leftrightarrow (\forall p, q \in A^*)(puq \in L \Leftrightarrow pvq \in L) \\ &\Leftrightarrow (\forall p, q \in A^*)(puq \in A^* \smallsetminus L \Leftrightarrow pvq \in A^* \smallsetminus L) \\ &\Leftrightarrow u\ \sigma_{A^*\smallsetminus L}\ v; \end{aligned}$$

Hence $\sigma_L = \sigma_{A^* \setminus L}$ and so $\mathrm{SynM}\, L = \mathrm{SynM}(A^* \setminus L)$.

b) Define a monoid homomorphism $\varphi : A^* \to (\mathrm{SynM}\, L) \times (\mathrm{SynM}\, K)$ by $u\varphi = (u\sigma_L^\natural, u\sigma_K^\natural)$. Let $S = L\sigma_L^\natural \times K\sigma_K^\natural \subseteq (\mathrm{SynM}\, L) \times (\mathrm{SynM}\, K)$. Then

$$\begin{aligned}
& u \in S\varphi^{-1} \\
\Rightarrow\ & u\varphi \in L\sigma_L^\natural \times K\sigma_K^\natural \\
\Rightarrow\ & (\exists v \in L, w \in K)\big((v\sigma_L^\natural, w\sigma_K^\natural) = (u\sigma_L^\natural, u\sigma_K^\natural)\big) \\
\Rightarrow\ & (\exists v \in L, w \in K)\big((v\sigma_L^\natural = u\sigma_L^\natural) \wedge (w\sigma_K^\natural = u\sigma_K^\natural)\big) \\
\Rightarrow\ & (u \in L\sigma_L^\natural(\sigma_L^\natural)^{-1}) \wedge (u \in K\sigma_K^\natural(\sigma_K^\natural)^{-1}) \\
\Rightarrow\ & (u \in L) \wedge (u \in K) \\
\Rightarrow\ & u \in L \cap K.
\end{aligned}$$

Hence $S\varphi^{-1} \subseteq L \cap K$. On the other hand,

$$\begin{aligned}
& u \in L \cap K \\
\Rightarrow\ & u \in L \wedge u \in K \\
\Rightarrow\ & u\varphi = (u\sigma_L^\natural, u\sigma_K^\natural) \in L\sigma_L^\natural \times K\sigma_K^\natural = S \\
\Rightarrow\ & u \in S\varphi^{-1},
\end{aligned}$$

so $S\varphi^{-1} \subseteq L \cap K$. Hence $S\varphi^{-1} = L \cap K$. Thus $\varphi : A^* \to (\mathrm{SynM}\, L) \times (\mathrm{SynM}\, K)$ recognizes $L \cap K$, and so $\mathrm{SynM}(L \cap K) \preccurlyeq (\mathrm{SynM}\, L) \times (\mathrm{SynM}\, K)$ by Proposition 9.6(a).

c) Define a monoid homomorphism $\varphi : A^* \to (\mathrm{SynM}\, L) \times (\mathrm{SynM}\, K)$ by $u\varphi = (u\sigma_L^\natural, u\sigma_K^\natural)$. Let $S = (L\sigma_L^\natural \times \mathrm{SynM}\, K) \cup (\mathrm{SynM}\, L \times K\sigma_K^\natural) \subseteq (\mathrm{SynM}\, L) \times (\mathrm{SynM}\, K)$. Then

$$\begin{aligned}
& u \in S\varphi^{-1} \\
\Rightarrow\ & u\varphi \in (L\sigma_L^\natural \times \mathrm{SynM}\, K) \cup (\mathrm{SynM}\, L \times K\sigma_K^\natural) \\
\Rightarrow\ & (\exists v \in L, w \in A^+)\big((v\sigma_L^\natural, w\sigma_K^\natural) = (u\sigma_L^\natural, u\sigma_K^\natural)\big) \\
& \qquad \vee (\exists v \in A^+, w \in K)\big((v\sigma_L^\natural, w\sigma_K^\natural) = (u\sigma_L^\natural, u\sigma_K^\natural)\big) \\
\Rightarrow\ & (\exists v \in L)(v\sigma_L^\natural = u\sigma_L^\natural) \vee (\exists w \in K)(w\sigma_K^\natural = u\sigma_K^\natural) \\
\Rightarrow\ & (u \in L\sigma_L^\natural(\sigma_L^\natural)^{-1}) \vee (u \in K\sigma_K^\natural(\sigma_K^\natural)^{-1}) \\
\Rightarrow\ & (u \in L) \vee (u \in K) \\
\Rightarrow\ & u \in L \cup K.
\end{aligned}$$

Hence $S\varphi^{-1} \subseteq L \cup K$. On the other hand,

$$\begin{aligned} & u \in L \cup K \\ \Rightarrow\ & u \in L \vee u \in K \\ \Rightarrow\ & u\varphi = (u\sigma_L^\natural, u\sigma_K^\natural) \in \\ & \qquad (L\sigma_L^\natural \times \operatorname{SynM} K) \cup (\operatorname{SynM} L \times K\sigma_K^\natural) = S \\ \Rightarrow\ & u \in S\varphi^{-1}, \end{aligned}$$

so $S\varphi^{-1} \subseteq L \cap K$. Hence $S\varphi^{-1} = L \cup K$. Thus $\varphi : A^+ \to (\operatorname{SynM} L) \times (\operatorname{SynM} K)$ recognizes $L \cup K$, and so $\operatorname{SynM}(L \cup K) \preccurlyeq (\operatorname{SynM} L) \times (\operatorname{SynM} K)$ by Proposition 9.6(a).

d) Let $S = L\sigma_L^\natural \subseteq \operatorname{SynM} L$. Let $s = a\sigma_L^\natural$. Define

$$s^{-1}S = \{\, x \in \operatorname{SynM} L : sx \in S \,\}.$$

Then $a^{-1}L = (s^{-1}S)\varphi^{-1}$ and so $\varphi : A^+ \to \operatorname{SynM} L$ recognizes $a^{-1}L$. Hence $\operatorname{SynM}(a^{-1}L) \preccurlyeq \operatorname{SynM} L$ by Proposition 9.6(a).

e) This is similar to part d).

f) The homomorphism $\varphi\sigma_L^\natural : B^+ \to \operatorname{SynM} L$ recognizes $L\varphi^{-1}$ since

$$L\varphi^{-1}\varphi\sigma_L^\natural(\varphi\sigma_L^\natural)^{-1} = L\varphi^{-1}\varphi\sigma_L^\natural(\sigma_L^\natural)^{-1}\varphi^{-1} = L\sigma_L^\natural(\sigma_L^\natural)^{-1}\varphi^{-1} = L\varphi^{-1}.$$

Hence $\operatorname{SynM}(L\varphi^{-1}) \preccurlyeq \operatorname{SynM} L$ by Proposition 9.6(a). 9.7

Essentially the same proofs yield the corresponding results for syntactic semigroups:

Properties of syntactic semigroups

PROPOSITION 9.8. *Let A and B be alphabets. For all $+$-languages K and L over A, for all $a \in A$, and for all homomorphisms $\varphi : B^+ \to A^+$:*

a) $\operatorname{SynS} L = \operatorname{SynS}(A^+ \smallsetminus L)$;

b) $\operatorname{SynS}(L \cap K) \preccurlyeq (\operatorname{SynS} L) \times (\operatorname{SynS} K)$;

c) $\operatorname{SynS}(L \cup K) \preccurlyeq (\operatorname{SynS} L) \times (\operatorname{SynS} K)$;

d) $\operatorname{SynS}(a^{-1}L) \preccurlyeq \operatorname{SynS} L$;

e) $\operatorname{SynS}(La^{-1}) \preccurlyeq \operatorname{SynS} L$;

f) $\operatorname{SynS}(L\varphi^{-1}) \preccurlyeq \operatorname{SynS} L$. 9.8

EILENBERG CORRESPONDENCE

The classes of languages that correspond to pseudovarieties are called 'varieties of rational languages'. However, the class of languages that are V-recognizable for some pseudovariety V is also dependent on the finite alphabet A as well. Thus we do not formally define

a 'variety of rational languages' as a class, but rather as a correspondence that associates finite alphabets to classes of languages. We also need to distinguish between $*$-languages and $+$-languages. To be precise, a *variety of rational $*$-languages* is formally defined to be a correspondence $\mathcal{V}$ that associates to each finite alphabet A a class of rational $*$-languages $\mathcal{V}(A^*)$ with the following properties:

Variety of rational $*$-languages

1) The class $\mathcal{V}(A^*)$ is a Boolean algebra. (That is, it is closed under union, intersection, and complement in A^*.)
2) For all $L \in \mathcal{V}(A^*)$ and $a \in A$, the right and left quotient languages
$$a^{-1}L = \{w \in A^* : aw \in L\} \quad \text{and} \quad La^{-1} = \{w \in A^* : wa \in L\}$$
are also in $\mathcal{V}(A^*)$
3) For all finite alphabets B, for all $*$-languages $L \in \mathcal{V}(B^*)$, and for all monoid homomorphisms $\varphi : A^* \to B^*$, we have $L\varphi^{-1} \in \mathcal{V}(A^*)$.

Similarly, a *variety of rational $+$-languages* is a correspondence $\mathcal{V}$ that associates to each finite alphabet A a class of rational languages $\mathcal{V}(A^+)$ with the following properties:

Variety of rational $+$-languages

1) The class $\mathcal{V}(A^+)$ is a Boolean algebra. (That is, it is closed under union, intersection, and complement in A^+.)
2) For all $L \in \mathcal{V}(A^+)$ and $a \in A$, the right and left quotient languages
$$a^{-1}L = \{w \in A^+ : aw \in L\} \quad \text{and} \quad La^{-1} = \{w \in A^+ : wa \in L\}$$
are also in $\mathcal{V}(A^+)$.
3) For all finite alphabets B, for all $+$-languages $L \in \mathcal{V}(B^+)$, and for all homomorphisms $\varphi : A^+ \to B^+$, we have $L\varphi^{-1} \in \mathcal{V}(A^+)$.

EXAMPLE 9.9. a) The correspondence $\mathcal{E}$ such that $\mathcal{E}(A^+) = \{\varnothing, A^+\}$ is a variety of rational $+$-languages. To see this, first note that each $\mathcal{E}(A^+)$ is clearly closed under union, intersection, and complement. Next, for any $a \in A$, we have $a^{-1}\varnothing = \varnothing a^{-1} = \varnothing \in \mathcal{E}(A^+)$ and $a^{-1}A^+ = A^+a^{-1} = A^+ \in \mathcal{E}(A^+)$, so $\mathcal{E}(A^+)$. Finally, for any homomorphism $\varphi : B^+ \to A^+$, we have $\varnothing\varphi^{-1} = \varnothing \in \mathcal{E}(B^+)$ and $A^+\varphi^{-1} = B^+ \in \mathcal{E}(B^+)$.

b) Let $\mathcal{M}$ be the correspondence that associates to each finite alphabet A the class of all $*$-languages over A. It is easy to see that $\mathcal{M}$ is a variety of rational $*$-languages.

c) A $+$-language L over an alphabet A is said to be *cofinite* if $A^+ \smallsetminus L$ is finite. Let $\mathcal{N}$ be the correspondence that associates to each finite alphabet A the class of all finite or cofinite languages over A. Then $\mathcal{N}$ is a variety of rational $+$-languages (see Exercise 9.4).

There is a natural correspondence, known as the *Eilenberg correspondence*, between varieties of rational $*$-languages and $\mathcal{M}$-pseudovarieties of monoids, and between varieties of rational $+$-languages and $\mathcal{S}$-pseudovarieties of semigroups. For varieties of rational $*$-languages and $\mathcal{M}$-pseudovarieties of monoids, the correspondence is defined as follows:

Eilenberg correspondence

- Let $\mathcal{V}$ be a variety of rational $*$-languages. The corresponding $\mathcal{M}$-pseudovariety of monoids V is generated by all monoids $\operatorname{SynM} L$ such that $L \in \mathcal{V}(A^*)$ for some finite alphabet A. That is, we have a map

$$\mathcal{V} \mapsto \mathsf{V}_{\mathcal{M}}\Big(\{ \operatorname{SynM} L : L \in \mathcal{V}(A^*) \text{ for some finite alphabet } A \} \Big). \tag{9.3}$$

- Let V be an $\mathcal{M}$-pseudovariety of monoids. The corresponding variety of rational $*$-languages $\mathcal{V}$ associates to each finite alphabet A the class of languages L such that $\operatorname{SynM} L \in \mathsf{V}$, or, equivalently, the class of languages L such that L is recognized by some monoid in V. That is, we have a map

$$\mathsf{V} \mapsto \mathcal{V}, \quad \text{where } \mathcal{V}(A^*) = \{ L \subseteq A^* : \operatorname{SynM} L \in \mathsf{V} \} \text{ for each finite alphabet } A. \tag{9.4}$$

The correspondence for $+$-varieties of rational languages and $\mathcal{S}$-pseudovarieties of semigroups is defined similarly:

- Let $\mathcal{V}$ be a variety of rational $+$-languages. The corresponding $\mathcal{S}$-pseudovariety of semigroups V is generated by all semigroups $\operatorname{SynS} L$ such that $L \in \mathcal{V}(A^+)$ for some finite alphabet A. That is, we have a map

$$\mathcal{V} \mapsto \mathsf{V}_{\mathcal{S}}\Big(\{ \operatorname{SynS} L : L \in \mathcal{V}(A^*) \text{ for some finite alphabet } A \} \Big). \tag{9.5}$$

- Let V be an $\mathcal{S}$-pseudovariety of semigroups. The corresponding variety of rational $+$-languages $\mathcal{V}$ associates to each finite alphabet A the class of languages L such that $\operatorname{SynS} L \in \mathsf{V}$, or, equivalently, the class of languages L such that L is recognized by some semigroup in V. That is, we have a map

$$\mathsf{V} \mapsto \mathcal{V}, \quad \text{where } \mathcal{V}(A^+) = \{ L \subseteq A^+ : \operatorname{SynS} L \in \mathsf{V} \} \text{ for each finite alphabet } A. \tag{9.6}$$

Eilenberg's theorem

EILENBERG'S THEOREM 9.10. *The maps* (9.3) *and* (9.4) *are mutually inverse, and the maps* (9.5) *and* (9.6) *are mutually inverse.*

Proof of 9.10. We will prove that (9.3) and (9.4) are mutually inverse; the other case is similar.

Let $\mathcal{V}$ be a variety of rational $*$-languages. Let V be the $\mathcal{M}$-pseudovariety of monoids associated to it by (9.3). Let $\mathcal{W}$ be the variety of rational $*$-languages associated to V by (9.4). We aim to show that $\mathcal{V}(A^*) = \mathcal{W}(A^*)$. for each finite alphabet A.

First, we prove that $\mathcal{V}(A^*) \subseteq \mathcal{W}(A^*)$. Let $L \in \mathcal{V}(A^*)$. Then $\operatorname{SynM} L \in \mathsf{V}$ by the definition of (9.3), and so $L \in \mathcal{W}(A^*)$ by the definition of (9.4). Hence $\mathcal{V}(A^*) \subseteq \mathcal{W}(A^*)$.

Next we prove that $\mathcal{W}(A^*) \subseteq \mathcal{V}(A^*)$. This part of the proof is more complicated. Let $L \in \mathcal{W}(A^*)$. Then $\operatorname{SynM} L \in \mathsf{V}$ by the definition of (9.4). Now, V is generated by

$$\mathcal{X} = \{ \operatorname{SynM} K : K \in \mathcal{V}(A^*) \text{ for some finite alphabet } A \};$$

that is, $\mathsf{V} = \mathbb{HSP}_{\text{fin}}\mathcal{X}$. Hence there exist alphabets A_i and $*$-languages $K_i \in \mathcal{V}(A_i^*)$ for $i = 1, \dots, n$ such that

$$\operatorname{SynM} L \preccurlyeq \prod_{i=1}^{n} \operatorname{SynM} K_i.$$

Let $U = \prod_{i=1}^{n} A_i^*$ and $T = \prod_{i=1}^{n} \operatorname{SynM} K_i$. Define a map

$$\gamma : U \to T, \qquad (w_1, \dots, w_n)\gamma = (w_1\sigma_{K_1}^\natural, \dots, w_n\sigma_{K_n}^\natural);$$

then γ is a surjective homomorphism because all of the maps $\sigma_{K_i}^\natural : A_i^* \to \operatorname{SynM} K_i$ are surjective homomorphisms. Since $\operatorname{SynM} L \preccurlyeq T$, the monoid T recognizes L; that is, there is a homomorphism $\varphi : A^* \to T$ and a subset M of T such that $L = M\varphi^{-1}$.

Define $\psi : A \to U$ by letting $a\psi$ be such that $a\psi\gamma = a\varphi$; since A^* is free on A, this map extends to a unique monoid homomorphism $\widehat{\psi} : A^* \to U$. Notice that $u\widehat{\psi}\gamma = u\varphi$ for $u \in A^*$ since φ and γ are monoid homomorphisms. For each $i = 1, \dots, n$, let $\psi_i : A^* \to A_i^*$ be such that

$$u\widehat{\psi} = (u\psi_1, \dots, u\psi_n)$$

and $\varphi_i : A^* \to \operatorname{SynM} K_i$ be such that

$$u\varphi = (u\varphi_1, \dots, u\varphi_n).$$

Then $\varphi_i = \psi_i\sigma_{K_i}^\natural$.

We have

$$L = M\varphi^{-1} = \bigcup_{m \in M} m\varphi^{-1}.$$

Since $\mathcal{V}(A^*)$ is a Boolean algebra, it is sufficient to show that $m\varphi^{-1} \in \mathcal{V}(A^*)$ for all $m \in M$. If $m = (s_1, \dots, s_n) \in M \subseteq T$, where $s_i \in \operatorname{SynM} K_i$, then

$$m\varphi^{-1} = \bigcap_{i=1}^{n} s_i\varphi_i^{-1}.$$

Again, since $\mathcal{V}(A^*)$ is a Boolean algebra, it is sufficient to show that $s_i\varphi_i^{-1} \in \mathcal{V}(A^*)$ for all $s_i \in \operatorname{SynM} K_i$ and $i = 1, \dots, n$.

Since $s_i\varphi_i^{-1} = s_i(\sigma_{K_i}^\natural)^{-1}\psi_i^{-1}$, the closure of $\mathcal{V}$ under homomorphism pre-images shows that it is sufficient to prove that $s_i(\sigma_{K_i}^\natural)^{-1} \in \mathcal{V}(A_i^*)$.

For $w \in A_i^*$, define

$$\begin{aligned} R_w &= \{ (p,q) : p,q \in A_i^*, pwq \in K_i \} \\ &= \{ (p,q) : p,q \in A_i^*, w \in p^{-1}K_i q^{-1} \}; \end{aligned}$$

then for any $u, v \in A^*$ we have $u \ \sigma_{K_i} \ v$ if and only if $R_u = R_v$. Hence $u\sigma_{K_i}^\natural(\sigma_{K_i}^\natural)^{-1}$, which is the σ_{K_i}-class of $u \in A_i^*$, is given by

$$u\sigma_{K_i}^\natural(\sigma_{K_i}^\natural)^{-1} = \bigcap_{(p,q)\in R_u} p^{-1}K_i q^{-1} \setminus \bigcup_{(p,q)\notin R_u} p^{-1}K_i q^{-1}. \tag{9.7}$$

By Proposition 9.3, there are only finitely many distinct sets $p^{-1}K_i q^{-1}$. Therefore the intersections and unions in (9.7) are finite. By repeated application of the closure of $\mathcal{V}(A_i^*)$ under left and right quotients, each of the sets $p^{-1}K_i q^{-1}$ lies in $\mathcal{V}(A_i^*)$. Since $\mathcal{V}(A^*)$ is closed under unions, intersections, and complements, $u\sigma_{K_i}^\natural(\sigma_{K_i}^\natural)^{-1}$ lies in $\mathcal{V}(A_i^*)$.

Finally, let u be such that $s_i = u\sigma_{K_i}^\natural$. Then $s_i(\sigma_{K_i}^\natural)^{-1} \in \mathcal{V}(A_i^*)$. This completes the proof. 9.10

It is important to notice that although Eilenberg's theorem shows that there is a one-to-one correspondence between $\mathcal{M}$-pseudovarieties of semigroups and varieties of rational $*$-languages, and between $\mathcal{S}$-pseudovarieties of monoids and varieties of rational $+$-languages, it does not actually give a concrete method for describing a variety of rational languages if we know a pseudovariety, or vice versa.

The following result is therefore an *instance* of the Eilenberg correspondence, but it is not a *consequence* of Theorem 9.10. In general, finding and proving instances of Eilenberg's correspondence can be difficult, although this particular result is straightforward.

THEOREM 9.11. *The Eilenberg correspondence associates the $\mathcal{S}$-pseudovariety of nilpotent semigroups* N *with the variety of finite or cofinite rational $+$-languages $\mathcal{N}$.*

Proof of 9.11. Let $S \in \mathsf{N}$, with $S^n = 0$ for all $x \in S$. Let A be a finite alphabet and suppose $\varphi : A^+ \to S$ recognizes a $+$-language L.

Suppose that $L\varphi$ contains 0_S. Then if $w \in A^*$ with $|w| \geqslant n$, then $w\varphi = 0_S$ and so $w \in L\varphi\varphi^{-1} = L$. Hence L contains all words with at least n letters and so is cofinite. Thus $L \in \mathcal{N}(A^+)$

Suppose that $L\varphi$ does not contain 0_S. Then if $w \in A^*$ with $|w| \geqslant n$, then $w \notin L$, since otherwise $L\varphi \ni w\varphi \in S^n = \{0\}$. Hence L contains only words with fewer than n symbols and so L is finite. Thus $L \in \mathcal{N}(A^+)$.

Thus if L is a $+$-language over A recognized by a semigroup in N, then $L \in \mathcal{N}(A^+)$.

On the other hand, let L be a $+$-language in $\mathcal{N}(A^+)$. So L is either finite or cofinite. Then there exists some $n \in \mathbb{N}$ such that either $L \cap I_n = \varnothing$ or

Variety of rational $*$-languages (class associated to A^*)	Symbol	$\mathcal{M}$-pseudo-variety	See also
$\{\varnothing, A^*\}$	$\mathcal{E}$	1	
Rational $*$-languages over A	$\mathcal{M}$	M	
Star-free $*$-languages over A	$\mathcal{SF}$	A	Th. 9.19

TABLE 9.3
Varieties of rational $*$-languages and $\mathcal{M}$-pseudovarieties related by the Eilenberg correspondence.

Variety of rational $+$-languages (class associated to A^+)	Symbol	$\mathcal{S}$-pseudo-variety	See also
$\{\varnothing, A^+\}$	$\mathcal{E}$	1	
Rational $+$-languages over A	$\mathcal{S}$	S	
Finite/cofinite $+$-langs over A	$\mathcal{N}$	N	Th. 9.11
$XA^* \cup Y$, where X, Y are finite $+$-languages	$\mathcal{K}$	K	Th. 9.12(a)
$A^*X \cup Y$, where X, Y are finite $+$-languages	$\mathcal{D}$	D	Th. 9.12(b)
$Z \cup \bigcup_{i=1}^{k} x_i A^* y_i$, where $\begin{cases} Z \text{ is a finite } +\text{-language,} \\ k \in \mathbb{N} \cup 0, \text{ and } x_i, y_i \in A^+ \end{cases}$	$\mathcal{L}1$	$\mathbb{L}$1	Th. 9.13
$Z \cup \bigcup_{a\in A} aA^*a \cup \bigcup_{i=1}^{k} a_i A^* a_i'$, where $\begin{cases} Z \subseteq A, \quad k \in \mathbb{N} \cup 0 \text{ and} \\ a_i, a_i' \in A \text{ with } a_i \neq a_i' \end{cases}$	$\mathcal{RB}$	RB	Exer. 9.5

TABLE 9.4
Varieties of rational $+$-languages and $\mathcal{S}$-pseudovarieties related by the Eilenberg correspondence.

$L \supseteq I_n$, where $I_n = \{w \in A^* : |w| \geqslant n\}$. Notice that I_n is an ideal of A^+, and that $S = A^+/I_n$ is a nilpotent semigroup with $S^n = 0_S$; thus $S \in \mathsf{N}$. Then the natural homomorphism $\rho_{I_n}^{\natural} : A^+ \to S$ recognizes L. 9.11

The remainder of this chapter is devoted to more involved results that, like Theorem 9.11, are instances of the Eilenberg correspondence. All such from this chapter are summarized in Tables 9.3 and 9.4.

A semigroup S is *left-trivial* if $es = e$ for all $e \in E(S)$ and $s \in S$, and *right-trivial* if $se = e$ for all $e \in E(S)$ and $s \in S$. The finite left-trivial semigroups form the $\mathcal{S}$-pseudovariety $\mathsf{K} = [\![x^\omega y = x^\omega]\!]_{\mathsf{S}}$, and the finite right-trivial semigroups form the $\mathcal{S}$-pseudovariety $\mathsf{D} = [\![yx^\omega = x^\omega]\!]_{\mathsf{S}}$

Let $\mathcal{K}$ be the correspondence where $\mathcal{K}(A^+)$ is the class of all $+$-languages of the form $XA^* \cup Y$, where X and Y are finite $+$-languages over A; and let $\mathcal{D}$ be correspondence where $\mathcal{D}(A^+)$ is the class of all $+$-languages of the form $A^*X \cup Y$, where X and Y are finite $+$-languages over A.

THEOREM 9.12. a) *$\mathcal{K}$ is a variety of $+$-languages and is associated by the Eilenberg correspondence to the $\mathcal{S}$-pseudovariety* K*;*

b) *$\mathcal{D}$ is a variety of $+$-languages and is associated by the Eilenberg correspondence to the $\mathcal{S}$-pseudovariety* D*.*

Proof of 9.12. We will prove part a); dual reasoning gives part b).

Let $L = XA^* \cup Y$, where X and Y are finite +-languages over A. Now, Y is finite and so lies in $\mathcal{N}(A^+)$. Hence $\mathrm{SynS}\, Y$ lies in N by Theorem 9.11. Hence $\mathrm{SynS}\, Y$ satisfies the $\mathcal{S}$-pseudoidentity $x^\omega y = x^\omega$ and so $\mathrm{SynS}\, Y \in \mathsf{K}$. Let n be the length of the longest word in X. Let $w \in A^+$ be such that $|w| = n$ and let $t \in A^*$. Then $uwv \in XA^* \Leftrightarrow uwtv \in XA^*$, so $w\ \sigma_{XA^*}\ wt$. So $st = s$ for all $s \in (\mathrm{SynS}\, XA^*)^n$ and $t \in \mathrm{SynS}\, XA^*$. In particular $et = e$ for all idempotents e, since $e^n = e$. Thus $\mathrm{SynS}\, XA^*$ satisfies the $\mathcal{S}$-pseudoidentity $x^\omega y = x^\omega$ and so $\mathrm{SynS}\, XA^* \in \mathsf{K}$. Finally, note that $\mathrm{SynS}\, L \preccurlyeq (\mathrm{SynS}\, XA^*) \times (\mathrm{SynS}\, Y)$ by by Proposition 9.6(b). Hence $\mathrm{SynS}\, L \in \mathbb{HSP}\mathsf{K} = \mathsf{K}$.

Now suppose that L is recognized by a semigroup $S \in \mathsf{K}$. Then there is a homomorphism $\varphi : A^+ \to S$ such that $L = L\varphi\varphi^{-1}$. Let $n = |S|$. Then, by Lemma 7.5, $S^n = SE(S)S = SE(S)$ since $ex = e$ for all $e \in E(S)$ and $x \in S$. Suppose that $wt \in L$ with $|w| = n$. Then $w\varphi \in S^n = SE(S)$ and so $w\varphi = se$ for some $s \in S$ and $e \in E(S)$. It follows that $(wt)\varphi = se(t\varphi) = se = w\varphi$ since $ex = e$ for all $x \in S$. Hence $w \in L$ and $wA^* \subseteq (seS)\varphi^{-1} = (se)\varphi^{-1} = w\varphi\varphi^{-1}$. Thus if $wt \in L$, where $|w| = n$, then $wA^* \subseteq L$. Hence $L = XA^* \cup Y$, where $X \subseteq A^n$ and Y is a set of words of length less than n, so $L \in \mathcal{K}(A^+)$. 9.12

Notice that a left- or right-trivial semigroup is also locally trivial: if S is left-trivial, then $es = e$ for all $s \in S$ and $e \in E(S)$, and hence $ese = e^2 = e$, which shows that S is locally trivial. Hence $\mathsf{K} \cup \mathsf{D} \subseteq \mathbb{L}\mathsf{1}$.

Let $\mathcal{L}1$ be the correspondence where $\mathcal{L}1(A^+)$ is the class of languages of the form

$$Z \cup \bigcup_{i=1}^{k} x_i A^* y_i, \tag{9.8}$$

where x_i and y_i are words A^+ and Z is a finite +-language.



Some texts define $\mathcal{L}1(A^+)$ to be the class of languages of the form $Z \cup XA^*Y$, where X, Y, and Z are finite +-languages, and claim that this is equivalent to (9.8). This is incorrect, because (for example) the language $a\{a,b\}^*a \cup b\{a,b\}^*b$ cannot be expressed in the form $Z \cup X\{a,b\}^*Y$.

THEOREM 9.13. *The Eilenberg correspondence associates the $\mathcal{S}$-pseudovariety $\mathbb{L}\mathsf{1}$ with the variety of +-languages $\mathcal{L}1$.*

Proof of 9.13. We will first of all show that the syntactic semigroups of wA^* and A^*w are in $\mathbb{L}\mathsf{1}$ for all $w \in A^+$. Since $wA^* \in \mathcal{K}(A^+)$, it follows by Theorem 9.12(a) that $\mathrm{SynS}(wA^*) \in \mathsf{K}$. Hence $\mathrm{SynS}(wA^*)$ is left-trivial and so locally trivial, and so $\mathrm{SynS}(wA^*) \in \mathbb{L}\mathsf{1}$. Similarly $\mathrm{SynS}(A^*w) \in \mathbb{L}\mathsf{1}$.

Let Z be a finite +-languages over A. Then

$$Z = \bigcup_{w \in Z} \{w\} = \bigcup_{w \in K} \Big(wA^* \setminus \bigcup_{a \in A} waA^* \Big).$$

So $\mathrm{SynS}\, Z \in \mathbb{L}1$ by Proposition 9.8 and since $\mathbb{L}1$ is closed under finitary direct products and division.

Furthermore, for any $x, y \in A^+$, we have

$$xA^*y = (xA^* \cap A^*y) \setminus \bigcup_{i=1}^{|x|+|y|} A^i.$$

So $\mathrm{SynS}(xA^*y) \in \mathbb{L}1$ by Proposition 9.8 again.

Therefore by Proposition 9.8,

$$\mathrm{SynS}\left(Z \cup \bigcup_{i=1}^{k} x_i A^* y_i \right) \in \mathbb{L}1.$$

On the other hand, let L be recognized by some semigroup S in $\mathbb{L}1$. Then S is locally trivial and there is a homomorphism $\varphi : A^+ \to S$ such that $L\varphi\varphi^{-1} = L$. Let $n = |S|$.

Let $e \in E(S)$ and $s \in S$. Then $(es)^2 = eses = es$ since $ese = e$ because S is locally trivial. Hence $es \in E(S)$. Similarly $se \in E(S)$. Thus $E(S)$ is an ideal, and so, by Lemma 7.5, $S^n = SE(S)S = E(S)$.

Let $w \in L$ be such that $|w| \geqslant 2n$. Then $w = xvy$ with $|x| = |y| = n$. Now, $S^n = E(S)$ by the previous paragraph, so $x\varphi, y\varphi \in S^n = E(S)$. Hence $w\varphi = (x\varphi)(v\varphi)(y\varphi) \in (x\varphi)S(y\varphi)$. Hence $xA^*y \subseteq w\varphi\varphi^{-1} \subseteq L$. Thus L is a finite union of languages of the form xA^*y (where $|x| = |y| = n$) and a finite set of words of length at most $2n$. Thus $L \in \mathcal{L}1(A^+)$. 9.13

COROLLARY 9.14. $\mathbb{L}1 = \mathsf{K} \sqcup \mathsf{D}$.

Proof of 9.14. Since $\mathsf{K} \cup \mathsf{D} \subseteq \mathbb{L}1$, it remains to show that $\mathbb{L}1 \subseteq \mathsf{K} \sqcup \mathsf{D}$.

Let $\mathcal{V}$ be the +-variety of rational languages associated to $\mathsf{K} \sqcup \mathsf{D}$ by the Eilenberg correspondence. Then $\mathcal{V}(A^+)$ contains the languages wA^* and A^*w and hence the Boolean algebra generated by these languages, namely $\mathcal{L}1(A^+)$. Therefore $\mathbb{L}1 \subseteq \mathsf{K} \sqcup \mathsf{D}$. 9.14

Notice that the proof of Corollary 9.14 essentially involves using the Eilenberg correspondence to convert a question about $\mathcal{S}$-pseudovarieties of semigroups into one about varieties of +-rational languages and back again. Although it would be possible to give a pure pseudovariety-theoretic proof of this result, the proof via the Eilenberg correspondence is much more straightforward.

SCHÜTZENBERGER'S THEOREM

The aim of this final section, and the capstone of the entire course, is Schützenberger's theorem, which shows that the star-free

rational languages are precisely those languages recognized by aperiodic monoids. Before embarking on the proof, we need to introduce a new concept.

Relational morphism

A *relational morphism* between two semigroups S and T is a relation $\varphi \subseteq S \times T$ such that

1) $x\varphi \neq \varnothing$ for all $x \in S$;
2) $(x\varphi)(y\varphi) \subseteq (xy)\varphi$ for all $x, y \in S$.

Notice that any homomorphism is also a relational morphism. We need to establish some basic properties of relational morphisms that we will use to prove Schützenberger's theorem. The first three lemmata follow immediately from this definition:

LEMMA 9.15. *A relational morphism $\varphi \subseteq S \times T$ between semigroups S and T is a subsemigroup of $S \times T$. The projection homomorphisms from $S \times T$ to S and T restricts to homomorphisms $\alpha : \varphi \to S$ and $\beta : \varphi \to T$ such that α is surjective and $\tau = \alpha^{-1}\beta$.* 9.15

LEMMA 9.16. *If $\varphi : S \to T$ is a surjective homomorphism from a semigroup S to a semigroup T, then $\varphi^{-1} \subseteq T \times S$ is a relational morphism between T and S.* 9.16

LEMMA 9.17. *If $\varphi \subseteq S \times T$ and $\psi \subseteq T \times U$ are relational morphisms between semigroups S and T, and between semigroups T and U, respectively, then $\varphi\psi \subseteq S \times U$ is a relational morphism between T and S.* 9.17

LEMMA 9.18. *Let $\varphi \subseteq S \times T$ be a relational morphism between finite semigroups S and T. Suppose that T is aperiodic and that for all $e \in E(T)$, the subsemigroup $e\varphi^{-1}$ is aperiodic. Then S is aperiodic.*

Proof of 9.18. Let $x \in S$. Since S is finite, $x^{k+m} = x^k$ for some $k, m \in \mathbb{N}$, and $H = \{x^k, x^{k+1}, \ldots, x^{k+m-1}\}$ is a subgroup of S. Let $\alpha : \varphi \to S$ and $\beta : \varphi \to T$ be as in Lemma 9.15, so that $\varphi = \alpha^{-1}\beta$. Then $H\alpha^{-1}$ is a subgroup of φ, and so $H\alpha^{-1}\beta = H\varphi$ is a subgroup of T. Since T is aperiodic $H\varphi$ is trivial by Proposition 7.4, $H\varphi = e$ for some idempotent e of T. By the hypothesis, $e\varphi^{-1} \supseteq H$ is aperiodic, and so $m = 1$ and $x^{k+1} = x^k$. Since $x \in S$ was arbitrary, this proves that S is aperiodic. 9.18

Schützenberger's theorem

SCHÜTZENBERGER'S THEOREM 9.19. *The Eilenberg correspondence associates the variety of star-free rational $*$-languages SF and the pseudovariety* A *of aperiodic monoids.*

Proof of 9.19. Let $\mathcal{A}$ be the $*$-variety of rational languages associated to A. We have to prove that $SF(A^*) = \mathcal{A}(A^*)$ for all finite alphabets A. So fix a finite alphabet A.

Part 1 [$SF(A^*) \subseteq \mathcal{A}(A^*)$]. The class of $*$-languages $SF(A^*)$ consists of the languages that can be obtained from the languages $\{a\}$ (for $a \in A$) and $\{\varepsilon\}$ using the Boolean operations and concatenation.

Let us therefore begin by showing that $\mathcal{A}(A^*)$ contains the languages $\{a\}$ (for $a \in A$) and $\{\varepsilon\}$. Let $a \in A$. Let $S = \{x, 0\}$ be a two-element null semigroup with all products equal to 0. Let $\psi : A \to S^1$ be the map with $a\psi = x$ and $b\psi = 0$ for all $b \in A \smallsetminus \{a\}$, and let $\psi^* : A^* \to S^1$ be the unique extension of ψ to a monoid homomorphism. It is easy to see that $\{a\} = \{x\}\psi^{-1} = \{a\}\psi\psi^{-1}$, thus $\{a\}$ is recognized by the monoid S^1. Clearly S^1 is an aperiodic monoid; thus $S^1 \in \mathsf{A}$. Hence $\{a\} \in \mathcal{A}(A^*)$ by (9.3). Finally, $\{\varepsilon\} = a^{-1}\{a\} \in \mathcal{A}(A^*)$ by the definition of a $*$-variety of rational languages.

Further, by the definition of a $*$-variety of rational languages, $\mathcal{A}(A^*)$ is a Boolean algebra and thus closed under the Boolean operations.

It therefore remains to show that $\mathcal{A}(A^*)$ is closed under concatenation. So let $K, L \in \mathcal{A}(A^*)$; we aim to prove that $KL \in \mathcal{A}(A^*)$. Then both $\mathrm{SynM}\, K$ and $\mathrm{SynM}\, L$ belong to A by (9.4). That is, $\mathrm{SynM}\, K$ and $\mathrm{SynM}\, L$ are aperiodic.

Consider the three syntactic monoid homomorphisms $\sigma_K^\natural : A^* \to \mathrm{SynM}\, K$, $\sigma_L^\natural : A^* \to \mathrm{SynM}\, L$, and $\sigma_{KL}^\natural : A^* \to \mathrm{SynM}(KL)$. Let $\eta = (\sigma_{KL}^\natural)^{-1}$. Since $\sigma_{KL}^\natural$ is surjective, $\eta \subseteq \mathrm{SynM}(KL) \times A^*$ is a relational morphism by Lemma 9.16. Let $\zeta : A^* \to \mathrm{SynM}\, K \times \mathrm{SynM}\, L$ be defined by $u\zeta = (u\sigma_K^\natural, u\sigma_L^\natural)$; clearly ζ is a homomorphism and thus a relational morphism. Let $\varphi = \eta\zeta$; then φ is a relational morphism between $\mathrm{SynM}(KL)$ and $(\mathrm{SynM}\, K) \times (\mathrm{SynM}\, L)$ by Lemma 9.17.

We want to use Lemma 9.18 and the relational morphism φ to show that $\mathrm{SynM}(KL)$ is aperiodic. Let $(e_1, e_2) \in E((\mathrm{SynM}\, K) \times (\mathrm{SynM}\, L))$. Let $m \in (e_1, e_2)\varphi^{-1}$. Then $m = g\eta$ for some $g \in (e_1, e_2)\zeta^{-1}$. Then $g^2\zeta = (e_1^2, e_2^2) = (e_1, e_2) = g\zeta$. Thus $(g^2, g) \in \ker\zeta \subseteq \ker\sigma_K^\natural = \sigma_K$. Similarly $(g^2, g) \in \sigma_L$.

Suppose $ug^2v \in KL$ for some $u, v \in A^*$. Then $ug^2v = xy$, for $x \in K$ and $y \in L$. Then, by equidivisibility, either there exists $p \in A^*$ such that $x = ugp$ and $py = gv$, or there exists $q \in A^*$ such that $xq = ug$ and $y = qgv$. Assume the former case; the latter is similar. Since $(g^2, g) \in \sigma_K$, we have $ug^2p \in K$, and so $ug^2py = ug^3v \in KL$. This shows that $ug^2v \in KL$ implies $ug^3v \in KL$.

Now suppose $ug^3v \in KL$ for some $u, v \in A^*$. Then $ug^3v = xy$, for $x \in K$ and $y \in L$. Then, by equidivisibility, either there exists $p \in A^*$ such that $x = ug^2p$ and $py = gv$, or there exists $q \in A^*$ such that $xq = ug^2$ and $y = qgv$. Assume the former case; the latter is similar. Since $(g^2, g) \in \sigma_K$, we have $ugp \in K$, and so $ugpy = ug^2v \in KL$. This shows that $ug^3v \in KL$ implies $ug^2v \in KL$.

Combining the last two paragraphs shows that $(g^3, g^2) \in \sigma_{KL}$, and so $m^3 = g^3\sigma_{KL}^\natural = g^2\sigma_{KL}^\natural = m^2$. Since m was an arbitrary element of $e\varphi^{-1}$, it follows that the subsemigroup $e\varphi^{-1}$ is aperiodic. Since both $\mathrm{SynM}\, K$ and $\mathrm{SynM}\, L$ are aperiodic, $(\mathrm{SynM}\, K) \times (\mathrm{SynM}\, L)$ is aperiodic. Hence, by Lemma 9.18, $\mathrm{SynM}(KL)$ is aperiodic. Thus $\mathrm{SynM}(KL) \in \mathsf{A}$, and so

$KL \in \mathcal{A}(A^*)$ by (9.3). So $\mathcal{A}(A^*)$ is closed under concatenation.

Thus $\mathcal{A}(A^*)$ contains every language in $\mathcal{SF}(A^*)$.

Part 2 [$\mathcal{A}(A^*) \subseteq \mathcal{SF}(A^*)$]. The aim is to prove that any $*$-language recognized by an aperiodic monoid M (and thus belonging to $\mathcal{A}(A^*)$) lies in $\mathcal{SF}(A^*)$. The strategy is to proceed by induction on $|M|$. For brevity, let $\Delta M = \{ N : N \preccurlyeq M \wedge N \neq M \}$. That is, ΔM is the class of monoids that strictly divide M. Let $A^*\Delta M$ be the class of $*$-languages over A recognized by some monoid in ΔM.

The base of the induction consists of the cases $|M| = 1$ and $|M| = 2$. First, suppose $|M| = 1$. Let L be a $*$-language over A recognized by $\varphi : A^* \to M$. Then either $L = \varnothing\varphi^{-1} = \varnothing$ or $L = M\varphi^{-1} = A^*$, and both these languages are in $\mathcal{SF}(A^*)$ by definition of a $*$-variety of rational languages.

Now suppose $|M| = 2$. Then M is the two-element semilattice $\{1, 0\}$ with $1 > 0$. [To see this, let $\{1, z\}$ be an aperiodic monoid. Then $11 = 1$, $1z = z1 = z$, and either $zz = 1$ or $zz = z$. But in the former case, we have a cyclic group, which is not aperiodic. Hence $zz = z$ and we have a commutative semigroup of idempotents.] Let L be a $*$-language recognized by $\varphi : A^* \to M$. Let $B = \{ a \in A : a\varphi = 0 \}$. Then

$$0\varphi^{-1} = \bigcup_{b \in B} A^* b A^*,$$
$$1\varphi^{-1} = A^* \setminus \bigcup_{b \in B} A^* b A^*.$$

Then $0\varphi^{-1} \in \mathcal{SF}(A^*)$, since $\mathcal{SF}(A^*)$ contains the languages A^* and $\{b\}$ for any $b \in B$ and is by definition closed under concatenation and union, and $1\varphi^{-1} \in \mathcal{SF}(A^*)$, since $\mathcal{SF}(A^*)$ is by definition closed under complementation. Since one of the four cases $L = \varnothing$, $L = 0\varphi^{-1}$, $L = 1\varphi^{-1}$, and $L = M\varphi^{-1} = 0\varphi^{-1} \cup 1\varphi^{-1}$ holds, and since $\mathcal{SF}(A^*)$ is closed under union, it follows that $L \in \mathcal{SF}(A^*)$.

We have completed the base of the induction; we turn now to the induction step. Let $|M| \geqslant 3$ and suppose that every language in $A^*\Delta M$ lies in $\mathcal{SF}(A^*)$; that is, $A^*\Delta M \subseteq \mathcal{SF}(A^*)$. We must prove that every language recognized by M lies in $\mathcal{SF}(A^*)$.

Let L be a $*$-language over A recognized by M. Then there exists a homomorphism $\varphi : A^* \to M$ and a subset P of M such that $L = P\varphi^{-1}$. If φ is not surjective, then L is recognized by the proper submonoid $\operatorname{im}\varphi$ of M and so, since $\operatorname{im}\varphi \in \Delta M$, by induction $L \in A^*\Delta M \subseteq \mathcal{SF}(A^*)$. So assume that φ is surjective. Furthermore, since

$$L = P\varphi^{-1} = \bigcup_{m \in P} m\varphi^{-1}$$

and $\mathcal{SF}(A^*)$ is by definition closed under union, it suffices to prove the case where $L = m\varphi^{-1}$.

Let

$$K = \bigcap \{ I : I \text{ is an ideal of } M \text{ and } |I| \geqslant 2 \}. \tag{9.9}$$

Then K is an ideal of M. For use later in the proof, we now establish some properties of K, considering separately the cases where M has a zero and where M does not have a zero:

- M has a zero. Let $D = K \setminus \{0\}$. Suppose D is non-empty. Then for any $x \in D$, we have $\{0, x\} \neq MxM \subseteq K$ (since K is an ideal and $x \neq 0$), and thus $MxM = K$ (since MxM is one of the ideals I in the intersection (9.9) and so $K \subseteq MxM$). So D is a single $\mathcal{J}$-class of M and so a single $\mathcal{D}$-class of M by Proposition 3.3. Furthermore, K is a 0-minimal ideal and so either 0-simple or null by Proposition 3.8(a). So either D is empty, or else D is a single $\mathcal{D}$-class and K is 0-simple or null.
- M does not have a zero. Then K is the kernel of M and so simple by Proposition 3.8(b). (Since if there were an ideal with only one element z, then $Sz = \{z\}$ and $zS = \{z\}$ and so z would be a zero.) Furthermore, $K^2 = K$. (Since if $K^2 \subsetneq K$, then K^2 would be an ideal of S strictly contained in K.)

We now consider separately the three cases where $m \notin K$, where m is the zero of M, and where m is not a zero of M (but M may or may not contain a zero):

a) $m \notin K$. Then there exists an ideal I of M with $|I| \geqslant 2$ such that $m \notin I$. Let $\rho_I = (I \times I) \cup \mathrm{id}_M$ be the Rees congruence. Then $m = \rho_I^\natural(\rho_I^\natural)^{-1}$ and hence $m\varphi^{-1} = m\rho_I^\natural(\rho_I^\natural)^{-1}\varphi^{-1} = (m\rho_I^\natural)(\varphi\rho_I^\natural)^{-1}$. Thus $m\varphi^{-1}$ is recognized by the homomorphism $\varphi\rho_I^\natural : A^* \to M/I$ and so is recognized by M/I. Since $|M/I| = |M| - |I| + 1 < |M|$ (since $|I| \geqslant 2|$), we have $M/I \in \Delta M$ and so by induction $m\varphi^{-1} \in SF(A^*)$.

b) M has a zero and $m = 0_M$. Let $C = \{ a \in A : a\varphi = 0 \}$. The first step is to prove that

$$0\varphi^{-1} = A^*CA^* \cup \bigcup_{(a,n,a') \in E} A^*a(n\varphi^{-1})a'A^*, \tag{9.10}$$

where

$$E = \{ (a, n, a') \in (A \setminus C) \times (M \setminus K) \times (A \setminus C) \\ : (a\varphi)n(a'\varphi) = 0 \wedge (a\varphi)n \neq 0 \wedge n(a'\varphi) \neq 0 \}.$$

First, notice that $(A^*CA^*)\varphi = M(C\varphi)M \subseteq \{0\}$. If $(a, n, a') \in E$, then $(A^*a(n\varphi^{-1})a'A^*)\varphi = M(a\varphi)n(a'\varphi)M = M0M = \{0\}$. This shows that the right-hand side of (9.10) is contained in the left-hand side.

Let $f \in 0\varphi^{-1} \setminus A^*CA^* = 0\varphi^{-1} \cap (A \setminus C)^*$. Since M has at least two elements, $1 \neq 0$ and so $f \neq \varepsilon$. Let g be the longest left factor

of f such that $g\varphi \neq 0$. (Such a left factor exists since $\varepsilon\varphi = 1$ and $f\varphi = 0$.) Then $f = ga'g'$, where $g\varphi \neq 0$ and $(ga')\varphi = 0$. Note that since $f \in (A \setminus C)^*$, we have $g \in (A \setminus C)^*$ and $a' \in A \setminus C$. Let h be the longest right factor of g such that $(ha')\varphi \neq 0$. (Such a right factor of g exists because $a'\varphi \neq 0$ and $(ga')\varphi = 0$.) Then $g = h'ah$, where $(ha')\varphi \neq 0$ and $(aha')\varphi = 0$. Note that since $g \in (A \setminus C)^*$, we have $h \in (A \setminus C)^*$ and $a \in A \setminus C$. Furthermore, since $g\varphi \neq 0$, we have $(ah)\varphi \neq 0$.

Let $n = h\varphi$. Suppose, with the aim of obtaining a contradiction, that $n \in K$. Since K is an ideal, $n(a'\varphi) \in K$. Since $n(a'\varphi) = (ha')\varphi \neq 0$, we have $n(a'\varphi) \in D$. Thus $n \mathcal{D} n(a'\varphi)$, and so, by Lemma 7.6(a), $n \mathcal{R} n(a'\varphi)$. Similarly, since $(a\varphi)n = (ah)\varphi \neq 0$, we have $(a\varphi)n \mathcal{L} n$. Hence, by Lemma 3.12, $(a\varphi)n(a'\varphi) \mathcal{L} n(a'\varphi)$ and therefore we have $(a\varphi)n(a'\varphi) \mathcal{D} n$. Thus $(a\varphi)n(a'\varphi)$ lies in the $\mathcal{D}$-class D, which contradicts the fact that $(aha')\varphi = 0$. Hence $n \notin K$, and thus $f = h'aha'g' \in A^*a(n\varphi^{-1})a'A^*$ with $(a, n, a') \in E$.

This shows that the left-hand side of (9.10) is contained in the right-hand side.

Since $n \notin K$, the reasoning in case a) shows that $n\varphi^{-1} \in A^*\Delta M$ and thus $n\varphi^{-1} \in SF(A^*)$. Since $SF(A^*)$ is closed under Boolean operations and concatenation, it follows from (9.10) that $m\varphi^{-1} = 0\varphi^{-1} \in SF(A^*)$.

c) $m \in K \setminus \{0\}$ (where $K \setminus \{0\} = K$ if M does not contain a zero). Now, $mM \subseteq K$ since K is an ideal. Hence all elements of $mM \setminus \{0\}$ are $\mathcal{D}$-related and hence $\mathcal{R}$-related by Lemma 7.6. Thus $R_m = mM \setminus \{0\}$. Similarly, $L_m = mM \setminus \{0\}$.

$$\begin{aligned} \{m\} &= H_m && \text{[by Proposition 7.4]} \\ &= R_m \cap L_m \\ &= (mM \setminus \{0\}) \cap (Mm \setminus \{0\}) \\ &= (mM \cap Mm) \setminus \{0\}. \end{aligned}$$

(When M does not contain a zero, this becomes $\{m\} = mM \cap Mm$.) Thus, since by case b) we already know that $0\varphi^{-1}$ is in $SF(A^*)$, it is sufficient to prove that $(mM)\varphi^{-1}$ and $(Mm)\varphi^{-1}$ are in $SF(A^*)$. We will prove $(mM)\varphi^{-1} \in SF(A^*)$; the other case is similar.

The first step is to prove that

$$(mM)\varphi^{-1} = 0\varphi^{-1} \cup \bigcup_{(n,a)\in F} (n\varphi^{-1})aA^*. \tag{9.11}$$

where

$$F = \{\, (n, a) \in (M - K) \times A : n(a\varphi) \in R_m \,\}.$$

(We formally let $0\varphi^{-1} = \varnothing$ if M does not contain a zero.)

If $(n, a) \in F$, then $(n\varphi^{-1})aA^* = R_m\varphi^{-1} \subseteq (mM)\varphi^{-1}$. Trivially, $0\varphi^{-1} \subseteq (mM)\varphi^{-1}$. Thus the right-hand side of (9.10) is contained in the left-hand side.

Let $f \in (mM)\varphi^{-1}$. If $f\varphi = 0$, then $f \in 0\varphi^{-1}$. So assume $f\varphi \neq 0$. Then $f\varphi \in mM \setminus \{0\} = R_m$. Since $1 \notin K$ (since M has at least two elements), $\varepsilon\varphi = 1 \notin R_m$. Let g be the longest left factor of f such that $g\varphi \notin R_m$. (Such a longest left factor exists since $\varepsilon\varphi \notin R_m$ and $f\varphi \in R_m$.) Hence $f = gag'$ where $g\varphi \notin R_m$ and $(ga)\varphi \in R_m$, where $g, g' \in A^*$ and $a \in A$.

Let $n = g\varphi$. Suppose, with the aim of obtaining a contradiction, that $n \in K$. Then $R_n = nM \setminus \{0\}$ and so $n(a\varphi) \in R_n$. But $n(a\varphi) \in R_m$, so $R_n = R_m$ and so $n \in R_m$, which contradicts $n = g\varphi \notin R_m$. Thus $n \notin K$, and so $(n, a) \in F$. Therefore $f \in (n\varphi^{-1})aA^*$

This shows that the left-hand side of (9.11) is contained in the right-hand side.

Thus, for any $(n, a) \in F$, we have $n \notin K$ and so $n\varphi^{-1} \in A^*\Delta M$ by case a), and thus $n\varphi^{-1} \in SF(A^*)$. Hence $(Mm)\varphi^{-1}$ is in $SF(A^*)$ by (9.11).

This completes the induction step and thus the proof. 9.19 Finis.

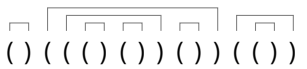
EXERCISES

[See pages 248–250 for the solutions.]

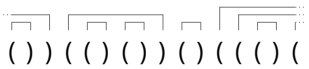
∗9.1 Prove that a language $L \subseteq A^*$ is rational if and only if SynM L is finite. [Hint: this is an easy consequence of results in this chapter.]

9.2 Let $A = \{a, b\}$. Let L be the language of words over A that contain at least one symbol a and at least one symbol b. (That is, $L = A^+ \setminus (\{a\}^+ \cup \{b\}^+)$.) Find a homomorphism $\varphi : A^+ \to S$ that recognizes L. [Hint: S can be taken to have 3 elements.]

∗9.3 A *Dyck word* is a string of balanced parentheses: that is, a word in $\{\,(\,,\,)\,\}^*$ where every opening parenthesis (matches a corresponding closing parentheses) to its right, and vice versa. For instance, Dyck word



is a Dyck word, but



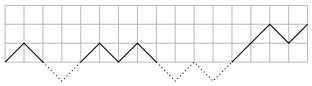
is not a Dyck word.

An equivalent characterization of Dyck words is the following: define a map $C : \{\,(\,,\,)\,\}^* \times \mathbb{N} \cup \{0\} \to \mathbb{Z}$ as follows: let $C(w_1 \cdots w_n, i)$ be the number of symbols (minus the number of symbols) in the prefix $w_1 \cdots w_i$. A word w is a Dyck word if and only if $C(w, |w|) = 0$ and

$C(w,i) \geqslant 0$ for all $i = 1, \ldots, |w|$. Consider the two example words above, with $C(w,i)$ plotted graphically:



If $w =$ ( ) ((( ) ( )) ( )) (( )) , then $C(w,|w|) = 0$ and $C(w,i) \geqslant 0$ for all i.



If $w =$ ( )) (( ) ( )) ( ) ((( ) (, then $C(w,|w|) \neq 0$ and $C(w,i) < 0$ for some i.

Let D be the language of Dyck words. Prove that SynM D is isomorphic to the bicyclic monoid.

✶9.4 Without using the Eilenberg correspondence, prove that the correspondence $\mathcal{N}$ in Example 9.9(c) (with $\mathcal{N}(A^+)$ being the class of finite or cofinite languages over A) is a variety of rational languages.

✶9.5 Let RB be the $\mathcal{S}$-pseudovariety of rectangular bands. Let $\mathcal{RB}$ be the variety of rational +-languages associated to RB by the Eilenberg correspondence. Prove that $\mathcal{RB}(A^+)$ is the class of all +-languages of the form

$$Z \cup \bigcup_{a \in Z} aA^*a \cup \bigcup_{i=1}^{n} a_i A^* a_i', \tag{9.12}$$

where $Z \subseteq A$ and $a_i, a_i' \in A$ with $a_i \neq a_i'$ for each i.

NOTES

For further reading on automata and rational languages, see Hopcroft & Ullman, *Introduction to Automata Theory, Languages, and Computation*, ch. 2, Lawson, *Finite Automata*, or Howie, *Automata and Languages*. ◆ Theorem 9.1 is due to Rabin & Scott, 'Finite automata and their decision problems'. ◆ Theorem 9.2 was first stated, in rather different terminology, in Kleene, 'Representation of events in nerve nets and finite automata'. ◆ The discussion of syntactic monoids and semigroups and Eilenberg's correspondence is based on Eilenberg, *Automata, Languages, and Machines* (Vol. B), ch. vii and Pin, 'Syntactic semigroups', §§ 2.2–3. ◆ The proof of Schützenberger's theorem is a blend of the original proof by Schützenberger, 'On finite monoids having only trivial subgroups' and its exposition in Pin, *Varieties of Formal Languages*, § 4.2 ◆ For further reading on the connection between semigroups and languages, see Pin, *Varieties of Formal Languages* or Pin, 'Mathematical Foundations of Automata Theory'.

Solutions to exercises

> A solution which does not prepare for the next round with some increased insight is hardly a solution at all.
>
> — R.W. Hamming, *The Art of Doing Science and Engineering*, p. 200.

EXERCISES FOR CHAPTER 1

[See pages 32–34 for the exercises.]

1.1 Let $x \in S$. Then $x = xe$ since e is a right identity, and $e = xe$ since e is a right zero. Hence $x = xe = e$. Thus e is the only element of S.

1.2 a) If S contains a zero, then $S^0 = S$ and there is nothing to prove. Otherwise $S^0 = S \cup \{0\}$. Then $x1 = x1 = x$ for all $x \in S$ since 1 is an identity for S, and $01 = 10 = 0$ by the definition of S^0. Hence 1 is an identity for S^0.

b) The reasoning is similar to part a).

1.3 Let S be left-cancellative and let $e \in S$ be an idempotent. Let $x \in S$. Since e is idempotent, $eex = ex$. Since S is left-cancellative, $ex = x$. Since $x \in S$ was arbitrary, this proves that e is a left identity for x.

Suppose now that S is cancellative and that $e, f \in S$ are idempotents. By the preceding paragraph and the symmetric result for right-cancellativity, e and f are left and right identities for S. By Proposition 1.3, $e = f$.

1.4 Let S be a right zero semigroup. Suppose $x, y, z \in S$ are such that $zx = zy$. Since S is a right zero semigroup, $zx = x$ and $zy = y$. Hence $x = zx = zy = y$. That is, $x = y$. So $zx = zy \Rightarrow x = y$ for all $x, y, z \in S$ and thus S is left-cancellative.

1.5 Let S be a finite cancellative semigroup. Let $x \in S$. Then x is periodic and so some power of x is an idempotent. By Exercise 1.3, this idempotent is an identity 1_S for S. Now let $y \in S$ be arbitrary. Then y^n is idempotent for some $n \in \mathbb{N}$. Again by Exercise 1.3, $y^n = 1_S$ and so y^{n-1} is a left and right inverse for y. Since $y \in S$ was arbitrary, S is a group.

1.6 Let $\rho \in \mathcal{B}_X$. Let $x, y \in X$. Then

$$\begin{aligned} & (x, y) \in \rho \circ \mathrm{id}_X \\ \Leftrightarrow\ & (\exists z \in X)\big((x, z) \in \rho \wedge (z, y) \in \mathrm{id}_X\big) && \text{[by definition of } \circ\text{]} \end{aligned}$$

$$\Leftrightarrow (\exists z \in X)\big((x,z) \in \rho \wedge (z = y)\big) \qquad \text{[by definition of } \mathrm{id}_X\text{]}$$
$$\Leftrightarrow (x,y) \in \rho.$$

So $\rho \circ \mathrm{id}_X = \rho$ and similarly $\mathrm{id}_X \circ \rho = \rho$. So id_X is the identity of $\mathcal{B}_X$.

The zero of $\mathcal{B}_X$ is the empty relation $\varnothing$. So see this, we must prove that $\rho \circ \varnothing = \varnothing \circ \rho = \varnothing$. So suppose, with the aim of obtaining a contradiction, that $\rho \circ \varnothing \neq \varnothing$. Then $(x,y) \in \rho \circ \varnothing$ for some $x, y \in X$. Then there exists $z \in X$ such that $(x,z) \in \rho$ and $(z,y) \in \varnothing$. But $(z,y) \in \varnothing$ is a contradiction. So $\rho \circ \varnothing = \varnothing$ and similarly $\varnothing \circ \rho = \varnothing$.

1.7 No. Let S be a non-trivial semigroup. Choose some element $x \in S$ and let $T = \{x^n : n \in \mathbb{N}\}$ be the subsemigroup generated by x. If T is finite (that is, if x is periodic), then some x^n is idempotent and so $\{x^n\}$ is a subsemigroup of S; furthermore, it must be a proper subsemigroup since S is non-trivial. If, on the other hand, T is infinite, then $\{x^{2n} : n \in \mathbb{N}\}$ is a proper subsemigroup of T and hence of S.

1.8 The easiest examples are infinite right or left zero semigroups, and the semigroups $(\mathbb{N}, \triangle)$ and $(\mathbb{Z}, \triangle)$ from Example 1.7(a)–(b).

1.9 The empty relation $\varnothing$ is a partial transformation. It is a zero for $\mathcal{B}_X$, so it is certainly a zero for $\mathcal{P}_X$. By Proposition 1.4, this is the unique left and right zero in $\mathcal{P}_X$.

Let us prove that the semigroup of transformations $\mathcal{T}_X$ contains exactly $|X|$ right zeros, namely the constant maps $\tau_x : X \to X$ defined by $y\tau_x = x$ for all $y \in X$. Each map τ_x is a right zero because for any $\sigma \in \mathcal{T}_X$, we have $y\sigma\tau_x = x$ for all $y \in X$, and so $\sigma\tau_x = \tau_x$. Suppose $\tau \in \mathcal{T}_X$ is a right zero. Then $\sigma\tau = \tau$ for all $\sigma \in \mathcal{T}_X$. In particular, this is true for all $\sigma \in \mathcal{S}_X$. Choose some $y \in X$ and let $x = y\tau$. Now let $z \in X$. Choose $\sigma \in \mathcal{S}_X$ with $z\sigma = y$. Then $z\tau = z\sigma\tau = y\tau = x$. Since $z \in X$ was arbitrary, we have $\tau = \tau_x$. Thus the right zeros in $\mathcal{T}_X$ are precisely the constant maps τ_x.

Suppose $\rho \in \mathcal{T}_X$ is a left zero. Then for all $x \in X$, we have $\rho = \rho\tau_x = \tau_x$ since ρ is a left zero and τ_x is a right zero. Hence $|X| = 1$ and so $\mathcal{T}_X$ is trivial (and so contains a zero). Hence if $|X| \geqslant 2$, then $\mathcal{T}_X$ cannot contain a left zero.

1.10 a) Define $\varphi : S \to \mathbb{P}S$ by $z\varphi = \{z\}$. Then

$$(z\varphi)(t\varphi) = \{z\}\{t\} = \{zt\} = (zt)\varphi.$$

and

$$z\varphi = t\varphi \Rightarrow \{z\} = \{t\} \Rightarrow z = t.$$

So φ is a monomorphism and so $\varphi : S \to \operatorname{im} S \subseteq \mathbb{P}S$ is an isomorphism.

b) For any $X \in \mathbb{P}S$, we have

$$X\varnothing = \{xy : x \in X \wedge y \in \varnothing\} = \varnothing$$

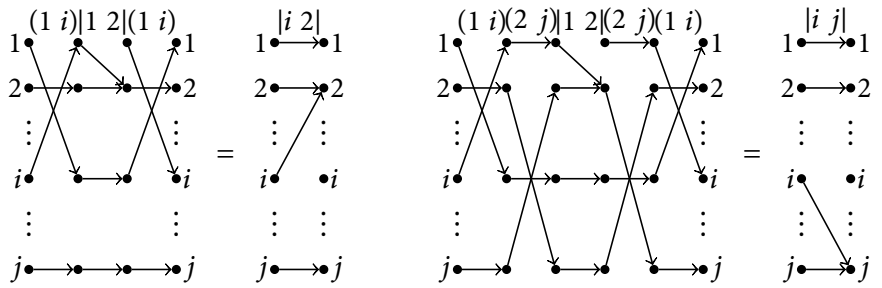


FIGURE S.3
Generating (a) $|2\ j|$ and (b) $|i\ j|$ using transpositions and $|1\ 2|$.

and similarly $\varnothing X = \varnothing$. So $\varnothing$ is a zero for $\mathbb{P}S$. If $X, Y \in (\mathbb{P}S) \setminus \{\varnothing\}$ then there exist $x \in X$ and $y \in Y$ and so $xy \in XY$ and hence $XY \neq \varnothing$. So $(\mathbb{P}S) \setminus \{\varnothing\}$ is a subsemigroup of $\mathbb{P}S$.

c) Suppose M is non-trivial. Let $x \in M \setminus \{1_M\}$. Let $N = M \setminus \{x\}$. Then $M \neq N \neq \varnothing$ but $MM = M$ and $MN = M$ since both M and N contain 1_M. Hence $(\mathbb{P}M) \setminus \{\varnothing\}$ is not cancellative.

On the other hand, suppose M is trivial. Then $\mathbb{P}M = \{\varnothing, M\}$. Hence $(\mathbb{P}M) \setminus \{\varnothing\}$ is trivial and thus cancellative.

d) Let S be a right zero semigroup. Let $X, Y \in (\mathbb{P}S) \setminus \{\varnothing\}$. Then

$$\begin{aligned} XY &= \{\, xy : x \in X \wedge y \in Y \,\} \\ &= \{\, y : x \in X \wedge y \in Y \,\} && \text{[since } y \text{ is a right zero]} \\ &= \{\, y : y \in Y \,\} \\ &= Y. \end{aligned}$$

So $(\mathbb{P}S) \setminus \{\varnothing\}$ is a right zero semigroup.

On the other hand, if $\mathbb{P}S$ is a right zero semigroup, so is its subsemigroup $\operatorname{im}\varphi \simeq S$, where φ is the monomorphism from part a). So S is a right zero semigroup.

1.11 a) To prove the four identities, we have to show that the transformations on each side act the same way on every element of X. For the first identity, let $i \geqslant 3$. Then:

$$\begin{aligned} 1(1\ i)|1\ 2|(1\ i) &= i|1\ 2|(1\ i) = i(1\ i) = 1 = 1|i\ 2|; \\ 2(1\ i)|1\ 2|(1\ i) &= 2|1\ 2|(1\ i) = 2(1\ i) = 2 = 2|i\ 2|; \\ i(1\ i)|1\ 2|(1\ i) &= 1|1\ 2|(1\ i) = 2(1\ i) = 2 = 2|i\ 2|; \\ x(1\ i)|1\ 2|(1\ i) &= x|1\ 2|(1\ i) = x(1\ i) = x = x|i\ 2| \\ &\qquad \text{for } x \in X \setminus \{1, 2, i\}. \end{aligned}$$

(Figure S.3(a) illustrates the first identity diagrammatically.) The second identity is proved similarly.

For the third identity, let $i, j \geqslant 3$ with $i \neq j$. Then:

$$\begin{aligned} 1(1\ i)(2\ j)|1\ 2|(2\ j)(1\ i) &= i(2\ j)|1\ 2|(2\ j)(1\ i) \\ &= i|1\ 2|(2\ j)(1\ i) = i(2\ j)(1\ i) = i(1\ i) = 1 = 1|i\ j|; \end{aligned}$$

$$
\begin{aligned}
&2(1\ i)(2\ j)|1\ 2|(2\ j)(1\ i) = 2(2\ j)|1\ 2|(2\ j)(1\ i)\\
&\qquad = j|1\ 2|(2\ j)(1\ i) = j(2\ j)(1\ i) = 2(1\ i) = 2 = 2|i\ j|;\\
&i(1\ i)(2\ j)|1\ 2|(2\ j)(1\ i) = 1(2\ j)|1\ 2|(2\ j)(1\ i)\\
&\qquad = 1|1\ 2|(2\ j)(1\ i) = 2(2\ j)(1\ i) = j(1\ i) = j = i|i\ j|;\\
&j(1\ i)(2\ j)|1\ 2|(2\ j)(1\ i) = j(2\ j)|1\ 2|(2\ j)(1\ i)\\
&\qquad = 2|1\ 2|(2\ j)(1\ i) = 2(2\ j)(1\ i) = j(1\ i) = j = j|i\ j|;\\
&x(1\ i)(2\ j)|1\ 2|(2\ j)(1\ i) = x(2\ j)|1\ 2|(2\ j)(1\ i)\\
&\qquad = x|1\ 2|(2\ j)(1\ i) = x(2\ j)(1\ i) = x(1\ i) = x = x|i\ j|\\
&\qquad\qquad\qquad\qquad \text{for } x \in X \setminus \{1, 2, i, j\}.
\end{aligned}
$$

(Figure S.3(b) illustrates the third identity diagrammatically.)

For the fourth identity, let $i \neq j$. Then:

$$
\begin{aligned}
&i(i\ j)|i\ j|(i\ j) = j|i\ j|(i\ j) = j(i\ j) = i = i|j\ i|;\\
&j(i\ j)|i\ j|(i\ j) = i|i\ j|(i\ j) = j(i\ j) = i = j|j\ i|;\\
&x(i\ j)|i\ j|(i\ j) = x|i\ j|(i\ j) = x(i\ j) = i = x|j\ i|\\
&\qquad\qquad\qquad\qquad \text{for } x \in X \setminus \{i, j\}.
\end{aligned}
$$

b) To prove that $|i\ j|\varphi' = \varphi$, we must show that both sides act the same way on every element of X. By the definition of φ',

$$
\begin{aligned}
&i|i\ j|\varphi' = j\varphi' = j\varphi,\\
&x|i\ j|\varphi' = x\varphi' = x\varphi \qquad\qquad \text{for } x \neq i.
\end{aligned}
$$

c) Since $\langle \tau, \zeta \rangle = \mathcal{S}_X$, we have $(i\ \ j) \in \langle \tau, \zeta, |1\ 2| \rangle$ for all $i, j \in X$. From part a), the first two identities show that $|i\ 2|$ and $|1\ \ j|$ are in $\langle \tau, \zeta, |1\ 2| \rangle$ for all $i, j \in X \setminus \{1, 2\}$. Combining this with the fourth identity shows that $|2\ \ j|$ and $|i\ 1|$ are in $\langle \tau, \zeta, |1\ 2| \rangle$. Together with the third identity, this shows that $|i\ \ j| \in \langle \tau, \zeta, |1\ 2| \rangle$ for all $i, j \in X$.

Now proceed by induction on $|X \setminus \operatorname{im} \varphi|$. If $|X \setminus \operatorname{im} \varphi| = 0$, then $\operatorname{im} \varphi = X$ and so φ is surjective and so (since X is finite) injective. Hence $\varphi \in \mathcal{S}_X = \langle \tau, \zeta \rangle \subseteq \langle \tau, \zeta, |1\ 2| \rangle$. So assume that $\psi \in \langle \tau, \zeta, |1\ 2| \rangle$ is true for all $\psi \in \mathcal{T}_X$ with $|X \setminus \operatorname{im} \psi| = k - 1 < n$. Let φ be such that $|X \setminus \operatorname{im} \varphi| = k$. Then by parts a) and b), we have $\varphi = |i\ \ j|\varphi' = (1\ i)(2\ j)|1\ 2|(2\ j)(1\ i)\varphi'$, where $\operatorname{im} \varphi' \subsetneq \operatorname{im} \varphi$. Hence $|X \setminus \operatorname{im} \varphi'| = k - 1$ and so $\varphi' \in \langle \tau, \zeta, |1\ 2| \rangle$. Hence $\varphi \in \langle \tau, \zeta, |1\ 2| \rangle$. By induction, $\mathcal{T}_X = \langle \tau, \zeta, |1\ 2| \rangle$.

1.12 Suppose x is right invertible. Then there exists $y \in S$ such that $xy = 1$. Since S is finite, $x^k = x^{k+m}$ for some $k, m \in \mathbb{N}$. So $1 = x^k y^k = x^{k+m} y^k = x^m = x^{m-1}x$ and so x^{m-1} is a left inverse for x. Similarly, if x is left invertible, it is right invertible.

1.13 a) Let $\rho \in \mathcal{T}_X$ be left-invertible. Then there exists $\sigma \in \mathcal{T}_X$ such that $\sigma \circ \rho = \mathrm{id}_X$. Let $x \in X$. Then $x(\sigma \circ \rho) = x$. So $(x\sigma)\rho = x$. So ρ is surjective.

Now let $\rho \in \mathcal{T}_X$ be surjective. Define $\sigma \in \mathcal{T}_X$ as follows. For each $x \in X$, choose $y \in X$ such that $y\rho = x$. (Such a y exists because ρ is surjective.) Define $x\sigma = y$. Clearly $\sigma \circ \rho = \mathrm{id}_X$ and so ρ is left-invertible.

b) Let $\rho \in \mathcal{T}_X$ be right-invertible. Then there exists $\sigma \in \mathcal{T}_X$ such that $\rho \circ \sigma = \mathrm{id}_X$. Then $x\rho = y\rho \Rightarrow (x\rho)\sigma = (y\rho)\sigma \Rightarrow x = y$ and so ρ is injective.

Now let $\rho \in \mathcal{T}_X$ be injective. Define $\sigma \in \mathcal{T}_X$ as follows. For $x \in \operatorname{im}\rho$, let $y \in X$ be the unique element such that $y\rho = x$. Define $x\sigma = y$. For $x \in X \smallsetminus \operatorname{im}\rho$, define $x\sigma$ arbitrarily. Clearly $\rho \circ \sigma = \mathrm{id}_X$ and so ρ is right-invertible.

1.14 a) By definition, $x \sqcap y \leqslant x$. So the least upper bound of $x \sqcap y$ and x (which is the definition of $(x \sqcap y) \sqcup x$) must be x itself. Dual reasoning gives $(x \sqcup y) \sqcap x = x$.

b) Assume that for all $p, q, r \in S$, we have $p \sqcap (q \sqcup r) = (p \sqcap q) \sqcup (p \sqcap r)$. (We have re-labelled variables to avoid confusion.) Then

$$
\begin{aligned}
&(x \sqcup y) \sqcap (x \sqcup z) \\
&= \big((x \sqcup y) \sqcap x\big) \sqcup \big((x \sqcup y) \sqcap z\big) \\
&\qquad\qquad \text{[by assumption, with } p = (x \sqcup y),\ q = x,\ r = z\text{]} \\
&= x \sqcup \big((x \sqcup y) \sqcap z\big) \qquad \text{[by part a)]} \\
&= x \sqcup \big((x \sqcap z) \sqcup (y \sqcap z)\big) \\
&\qquad\qquad \text{[by assumption, with } p = z,\ q = x,\ r = y\text{]} \\
&= \big(x \sqcup (x \sqcap z)\big) \sqcup (y \sqcap z) \qquad \text{[by associativity of } \sqcup\text{]} \\
&= x \sqcup (y \sqcap z). \qquad \text{[by part a)]}
\end{aligned}
$$

The other direction is similar.

1.15 There are many examples. For instance, let S be any non-trivial monoid, let $T = S^0$, and define $\varphi : S \to T$ by $x\varphi = 0$ for all $x \in S$. It is easy to see that φ is a homomorphism, but $1_S\varphi = 0 \neq 1_{S^0}$.

1.16 a) Let φ be a monomorphism (that is, an injective homomorphism), and let $\psi_1, \psi_2 : U \to S$ be such that $\psi_1 \circ \varphi = \psi_2 \circ \varphi$. Let $x \in U$. Then $x\psi_1\varphi = x\psi_2\varphi$ and so $x\psi_1 = x\psi_2$ since φ is injective. Since this is true for all $x \in U$, it follows that $\psi_1 = \psi_2$. This proves that φ is a categorical monomorphism.

Now let φ be a categorical monomorphism. Suppose, with the aim of obtaining a contradiction, that φ is not injective. Then there exist $x, y \in S$ with $x \neq y$ such that $x\varphi = y\varphi$. Define maps $\psi_1, \psi_2 : \mathbb{N} \to S$ by $n\psi_1 = x^n$ and $n\psi_2 = y^n$. It is easy to see that ψ_1 and ψ_2 are homomorphisms. Then for any $n \in \mathbb{N}$,

$$n\psi_1\varphi = x^n\varphi = (x\varphi)^n = (y\varphi)^n = y^n\varphi = n\psi_2\varphi,$$

and so $\psi_1 \circ \varphi = \psi_2 \circ \varphi$. Hence $\psi_1 = \psi_2$ by (1.16), which contradicts $1\psi_1 = x \neq y = 1\psi_2$ and so proves that φ is a monomorphism.

b) i) Let φ be a surjective homomorphism. Suppose $\psi_1, \psi_2 : T \to U$ are such that $\varphi \circ \psi_1 = \varphi \circ \psi_2$. Let $x \in T$. Then since φ is surjective, there exists $y \in S$ with $y\varphi = x$. Thus

$$x\psi_1 = y\varphi\psi_1 = y\varphi\psi_2 = x\psi_2.$$

Since this holds for all $x \in T$, it follows that $\psi_1 = \psi_2$. This proves that φ is a categorical epimorphism.

ii) Let $\psi_1, \psi_2 : \mathbb{Z} \to U$ be such that $\psi_1 \neq \psi_2$ (which is the negation of the right-hand side of (1.17)). Then there exists $n \in \mathbb{Z}$ such that $n\psi_1 \neq n\psi_2$. Either n or $-n$ lies in $\operatorname{im} \iota$, and so either $n\iota\psi_1 \neq n\iota\psi_2$ or $(-n)\iota\psi_1 \neq (-n)\iota\psi_2$, and thus $\iota\circ\psi_1 \neq \iota\circ\psi_2$ (which is the negation of the left-hand side of (1.17) with $\varphi = \iota$). Thus ι is a categorical epimorphism.

1.17 Suppose S is a right zero semigroup. Let $x, y \in S$. Then $\rho_x = \rho_y \Rightarrow z\rho_x = z\rho_y \Rightarrow zx = zy \Rightarrow x = y$ and so the map $x \mapsto \rho_x$ is injective.

Suppose now that S is a left zero semigroup. Let $x, y \in S$ with $x \neq y$. Then $zx = zy$ for all $z \in S$. Hence $z\rho_x = z\rho_y$ for all $z \in S$, and so $\rho_x = \rho_y$. Thus $x \mapsto \rho_x$ is not injective.

1.18 For each $y \in Y$, let T_y be a copy of T, and define a map $\varphi_y : Y \to T_y$

$$x\varphi_y = \begin{cases} e & \text{if } x \geqslant y, \\ z & \text{otherwise.} \end{cases}$$

Let $x, x', y \in Y$. Then

$$\begin{aligned}
(x\varphi_y) \sqcap (x'\varphi_y) &= e \sqcap e = e = (x \sqcap x')\varphi_y && \text{if } x, x' \geqslant y; \\
(x\varphi_y) \sqcap (x'\varphi_y) &= e \sqcap z = z = (x \sqcap x')\varphi_y && \text{if } x \geqslant y, x' \ngeq y; \\
(x\varphi_y) \sqcap (x'\varphi_y) &= z \sqcap e = z = (x \sqcap x')\varphi_y && \text{if } x \ngeq y, x' \geqslant y; \\
(x\varphi_y) \sqcap (x'\varphi_y) &= z \sqcap z = z = (x \sqcap x')\varphi_y && \text{if } x, x' \ngeq y.
\end{aligned}$$

So φ_y is a homomorphism. It is clearly surjective. Now,

$$\begin{aligned}
&(\forall y \in Y)(x\varphi_y = x'\varphi_y) \\
&\Rightarrow (x\varphi_x = x'\varphi_x) \wedge (x\varphi_{x'} = x'\varphi_{x'}) \\
&\Rightarrow (e = x'\varphi_x) \wedge (x\varphi_{x'} = e) \\
&\Rightarrow (x' \geqslant x) \wedge (x \geqslant x') \\
&\Rightarrow x = x'.
\end{aligned}$$

So the collection of surjective homomorphisms $\{\varphi_y : Y \to T_y : y \in Y\}$ separates elements of Y, and so Y is a subdirect product of $\{T_y : y \in Y\}$.

1.19 Define a homomorphism $\varphi : S/I \to S/J$ by $[x]_I\varphi = [x]_J$. Since $I \subseteq J$, the homomorphism φ is well defined. Its image is clearly S/J. Now, $([x]_I, [y]_I) \in \ker\varphi \Leftrightarrow [x]_J = [y]_J \Leftrightarrow x, y \in J \Leftrightarrow [x]_I, [y]_I \in J/I$. Hence, by Theorem 1.24, $S/J \simeq (S/I)/\ker\varphi \simeq (S/I)/(J/I)$.

1.20 Notice that $IJ \subseteq IS \cap SJ \subseteq I \cap J$, so $I \cap J \neq \varnothing$. Furthermore, $S(I \cap J)S \subseteq SIS \cap SJS \subseteq I \cap J$, since I and J are ideals; thus $I \cap J$ is an ideal. Similarly, $S(I \cup J)S \subseteq SIS \cup SJS \subseteq I \cup J$ and so $I \cup J$ is an ideal.

Define a homomorphism $\varphi : I \to (I \cup J)/J$ by $x\varphi = [x]_J$. Let $[y]_J \in (I \cup J)/J$. If $y \in J$ then let $z \in I \cap J$ and notice that $z\varphi = [z]_J = [y]_J$; if $y \notin J$ then $y \in I$ and $y\varphi = [y]_J$. Hence $\operatorname{im}\varphi$ is $(I \cup J)/J$. Now for any $x, y \in I$, we have $(x, y) \in \ker\varphi \Leftrightarrow [x]_J = [y]_J \Leftrightarrow x, y \in J$. Hence $(I \cup J)/J \simeq I/(I \cap J)$ by Theorem 1.24.

1.21 Suppose first that $T = G \cup \{0_S\}$ and let $t \in T \setminus \{0_S\} = G$. Then $tG = Gt = G$ by Lemma 1.9 and $t0_S = 0_S t = 0_S$, so $tT = Tt = T$.

Conversely, suppose that $tT = Tt = T$ for all $t \in T \setminus \{0_S\}$. Let $G = T \setminus \{0_S\}$. By assumption, T contains at least one element other than 0_S, so $G \neq \varnothing$. For any $s, t \in T$, we have $s, t \in sT = T$, so T is a subsemigroup.

Suppose, with the aim of obtaining a contradiction, that there exist $g, h \in G$ with $gh = 0_S$. Then

$$T = gT \subseteq TT = (Tg)(hT) = T(gh)T = T0_S T = \{0_S\},$$

contradicting $G \neq \varnothing$. So for all $g, h \in G$, we have $gh \in G$. Since $g0_S = 0_S g = 0_S$, it follows that $gG = Gg = G$ for all $g \in G$. Hence, by Lemma 1.9, G is a subgroup of S.

EXERCISES FOR CHAPTER 2

[See pages 51–53 for the exercises.]

2.1 a) Let G be a group and suppose that $x, y, z, t \in G$ are such that $xy = zt$. Then we can take $p = z^{-1}x = ty^{-1}$, and it follows that $x = zp$ and $t = py$.

b) Suppose $x, y, z, t \in A^*$ are such that $xy = zt$. Let $xy = zt = a_1 \cdots a_n$, where $a_i \in A$. Then, by the definition of multiplication in A^*, we have

$$x = a_1 \cdots a_k, \quad y = a_{k+1} \cdots a_n, \quad z = a_1 \cdots a_\ell, \quad t = a_{\ell+1} \cdots a_n,$$

for some $0 \leqslant k, \ell \leqslant n+1$. (We allow k and ℓ to take the values 0 and $n+1$ and formally take subwords $a_i \cdots a_j$ where $j < i$ to mean the empty word ε.) If $k \leqslant \ell$, then the situation is as follows:

$$xy = zt = \underbrace{\overbrace{a_1 \cdots a_k}^{x}\,\overbrace{a_{k+1} \cdots a_\ell}^{z}}_{z}\,\underbrace{\overbrace{a_{\ell+1} \cdots a_n}^{y}}_{t}$$

and thus we let $q = a_{k+1} \cdots a_\ell$; then $z = xq$ and $y = qt$. On the other hand, if $k \geqslant \ell$, let $p = a_{\ell+1} \cdots a_k$; then $x = zp$ and $t = py$.

2.2 If $u = w^i$ and $v = w^j$, then $uv = w^{i+j} = vu$.

In the other direction, suppose that $uv = vu$. First note that if $u = \varepsilon$, then we can take $w = v$, so that $u = w^0$ and $v = w^1$; similar reasoning holds when $v = \varepsilon$. So assume henceforth that neither u nor v is the empty word. Now proceed by induction on $|uv|$. If $|uv| \leqslant 2$ and $uv = vu$, then since neither u nor v is ε, it follows that u and v both have length 1, so $u = v$. So assume the result holds for $|uv| < k$ and suppose $uv = vu$. By Exercise 2.1, there exists $p \in A^*$ such that $u = vp$ and $u = pv$ (or there exists $q \in A^*$ such that $v = uq$ and $v = qu$, which says the same thing). If $p = \varepsilon$ then $u = v$. Otherwise, since $vp = pv$, the induction hypothesis shows that $v = w^j$ and $p = w^i$ for some $w \in A^*$ and $i, j \in \mathbb{N}$. Thus $u = w^{i+j}$. Hence, by induction, the result holds for all $u, v \in A^*$.

2.3 a) Suppose $uv = vw$. If $|v| = 0$, then let $s = \varepsilon$, $t = u$, $k = 0$. Since $v = \varepsilon$ and $u = w$, we have $u = st$, $v = (st)^k s$, $w = ts$. So suppose the result holds for $|v| < k$. Then if $|v| = k$, by equidivisibility we have either $u = vp$ and $pv = w$ for some $p \in A^*$ or $uq = v$ and $v = qw$ for some $q \in A^*$. In the former case, let $s = v$, $t = p$, and $k = 0$; then $u = st$, $v = (st)^k s$, and $w = ts$. In the latter case, first note that if $|q| = 0$ we have $uq = qw$, with $|q| < |v|$. By the induction hypothesis, $u = st$, $t = (st)^k s$, and $w = ts$ for some $s, t \in A^*$ and $k \in \mathbb{N} \cup \{0\}$. Then $v = uq = (st)^{k+1} s$. This proves the induction step.

b) Let k be maximal such that $v = u^k s$ for some $s \in A^*$. Then $u^{k+1} s = uv = vw = u^k sw$ and so by cancellativity $us = sw$. So by equidivisibility, either s is a left factor of u or u is left factor of s. But the latter contradicts the maximality of k. Hence $u = st$ for some $t \in A^*$. Hence $v = (st)^k s$ and so $(st)^{k+1} s = uv = vw = (st)^k sw$ and so by cancellativity $w = ts$.

2.4 First, notice that if $\langle u, v \rangle$ is free, then every element of $\langle u, v \rangle$ has a unique representation as a a product of elements of $\{u, v\}$; hence $uv \neq vu$.

So suppose $\langle u, v \rangle$ is not free. Without loss of generality, assume $|u| \geqslant |v|$ and let $u = v^k z$, where $k \in \mathbb{N} \cup \{0\}$ is maximal and $z \in A^*$. Then there are two distinct products $x_1 \cdots x_m$ and $y_1 \cdots y_n$ (where $x_i, y_i \in \{u, v\}$) such that $x_1 \cdots x_m = y_1 \cdots y_n$. By cancellativity, assume $x_1 \neq y_1$. Interchanging the two products if necessary, assume $x_1 = u$ and $y_1 = v$. Let $\ell \in \mathbb{N}$ be maximal such that $y_1 = y_2 = \ldots = y_\ell = v$. Then $v^k z x_2 \cdots x_m = v^\ell y_{\ell+1} \cdots y_n$. By cancellativity, $z x_2 \cdots x_m = v^{\ell-k} y_{\ell+1} \cdots y_n$. By equidivisibility, either $z = vp$ and $p x_2 \cdots x_m = v^{\ell-k} y_{\ell+1} \cdots y_n$ for some $p \in A^*$, or $v = zq$ and $q v^{\ell-k} y_{\ell+1} \cdots y_n$. The former case is impossible since k is maximal; thus the latter case holds. So $u = (zq)^k z$. Repeat this reasoning but focusing on u_m and v_n shows that v is a right factor of u. But since $u = (zq)^k z$ and $|v| =$

$|z| + |q| = |qz|$, we conclude that $v = qz$. Hence $zq = v = qz$, and so $uv = (qz)^k zqz = (zq)^k zqz = zq(zq)^k z = zq(qz)^k z = vu$.

2.5 Suppose that $p_1 \dots p_k = q_1 \dots q_\ell$, where $p_i, q_i \in X$. Suppose, with the aim of obtaining a contradiction, that $k \neq \ell$. Without loss of generality, assume $k < \ell$. Let $r \in X \setminus \{q_{k+1}\}$; such an element r exists since $|X| \geqslant 2$. Now, $p_1 \cdots p_k r q_1 \cdots q_\ell = q_1 \cdots q_\ell r p_1 \cdots p_k$. Both products have length $k + \ell + 1$ and so their corresponding terms are equal by the supposition. In particular, $r = q_{k+1}$, which contradicts the choice of r. Hence $k = \ell$, and so by the supposition $p_i = q_i$ for all i. Since S is generated by X, this proves that S is free with basis X.

2.6 a) Define $\varphi : A \to M$ by $a_i \mapsto \{x_i\}$. Since $(a_i a_j)\varphi = \{x_i\} \cup \{x_j\} = \{x_i, x_j\} = \{x_j\} \cup \{x_i\} = (a_j a_i)\varphi$ and $(a_i^2)\varphi = \{x_i\} \cup \{x_i\} = \{x_i\} = a_i\varphi$, the monoid M satisfies the defining relations in ρ with respect to φ.

b) Let $w \in A^*$. We can find a sequence of elementary transition from w to a word $a_1^{e_1} a_2^{e_2} \cdots a_n^{e_n} \in N$, where each $e_i \leqslant 1$ as follows. First we use the defining relations $(a_i a_j, a_j a_i)$ to find a sequence from w to a word $a_1^{e_1} a_2^{e_2} \cdots a_n^{e_n}$, where each $e_i \in \mathbb{N} \cup \{0\}$. Then we use the defining relations (a_i^2, a_i) to find a sequence from this word to one where each $e_i \leqslant 1$.

c) Let $a_1^{e_1} a_2^{e_2} \cdots a_n^{e_n}, a_1^{c_1} a_2^{c_2} \cdots a_n^{c_n} \in N$ so that $e_i, c_i \leqslant 1$. Then:

$$\begin{aligned} & (a_1^{e_1} a_2^{e_2} \cdots a_n^{e_n})\varphi^* = (a_1^{c_1} a_2^{c_2} \cdots a_n^{c_n})\varphi^* \\ \Rightarrow\ & \{ x_i : e_i = 1 \} = \{ x_i : c_i = 1 \} \\ \Rightarrow\ & (\forall i)(e_i = c_i) \\ \Rightarrow\ & a_1^{e_1} a_2^{e_2} \cdots a_n^{e_n} = a_1^{c_1} a_2^{c_2} \cdots a_n^{c_n}. \end{aligned}$$

Hence $\varphi^*|_N$ is injective.

2.7 We apply Method 2.9. For brevity, let $A = \{a, b\}$ and $\rho = \{(aba, \varepsilon)\}$. Let $\varphi : A \to \mathbb{Z}$ be defined by $a\varphi = 1$ and $b\varphi = -2$. Then $\mathbb{Z}$ satisfies the defining relation in ρ since $(aba)\varphi^* = 1 - 2 + 1 = 0 = \varepsilon\varphi^*$. [Recall that 0 is the identity of $\mathbb{Z}$ under addition.] Let

$$N = \{ a^i : i \in \mathbb{N} \cup \{0\} \} \cup \{ b^i : i \in \mathbb{N} \} \cup \{ ab^i : i \in \mathbb{N} \}.$$

Now, there are sequences of elementary transitions

$$ab \leftrightarrow_\rho ababa \leftrightarrow_\rho ba$$

and

$$aab \leftrightarrow_\rho aababa \leftrightarrow_\rho aba \leftrightarrow_\rho \varepsilon.$$

Thus we can first of all transform any word in A^+ by applying defining relations to replace subwords ba by ab, which ultimately yields a word

of the form $a^i b^j$. Then we can replace subwords aab by ε, which must ultimately yield a word consisting either entirely of symbols a, entirely of symbols b, or by a single symbol a followed by symbols b; that is, a word in N. Finally note that

$$\begin{aligned} a^i\varphi^* &= i && \text{for } i \in \mathbb{N} \cup \{0\}, \\ b^i\varphi^* &= -2i && \text{for } i \in \mathbb{N}, \\ (ab^i)\varphi^* &= -2i + 1 && \text{for } i \in \mathbb{N}. \end{aligned}$$

It is now easy to see that $\varphi^*|_N$ is injective. Hence $\mathrm{Mon}\langle A \mid \rho\rangle$ defines $(\mathbb{Z}, +)$

2.8 Deleting a subword abc is an elementary ρ-transition, and so does not alter the element represented. Thus given any word $w \in A^*$, one can obtain a word $\widehat{w} \in N$ with $w =_M \widehat{w}$ by deleting subwords abc. Thus every element of M has at least one representative in N; it remains to prove uniqueness.

So suppose some element of M has two representatives $u, v \in N$ with $u \neq v$. Since $u =_M v$, there is a sequence of elementary ρ-transitions

$$u = w_0 \leftrightarrow_\rho w_1 \leftrightarrow_\rho \ldots \leftrightarrow_\rho w_n = v.$$

Consider the collection of such sequences with the maximum length of an intermediate word w_i being minimal, and choose and fix such a sequence where the fewest words w_i have this maximum length. Note that $n > 0$ since $u \neq v$. Consider some intermediate word w_i of this maximum length. Note that $i \neq 0$ and $i \neq n$, since the words w_0 and w_n do not contain subwords abc, and so the words w_1 and w_n must be obtained by inserting subwords abc into w_0 and w_n respectively, and so $|w_1| > |w_0|$ and $|w_{n-1}| > |w_n|$. So there are words w_{i-1} and w_{i+1}, and these are obtained from w by applying the defining relation (abc, ε). Because w_i has maximum length among the intermediate words, w_{i-1} and w_{i+1} must both be obtained by deleting subwords abc from w_i. Now, they cannot be obtained by deleting the *same* subword abc, for then

$$u = w_0 \leftrightarrow_\rho w_1 \leftrightarrow_\rho \ldots \leftrightarrow_\rho w_{i-1} = w_{i+1} \leftrightarrow_\rho \ldots \leftrightarrow_\rho w_n = v.$$

would be a sequence of elementary ρ-transitions from u to v where the number of intermediate words of maximum length is smaller, or (if no other intermediate word had length $|w_i|$) a smaller maximum length of intermediate words; in either case, this is a contradiction. Hence w_{i-1} and w_{i+1} are obtained by deleting different subwords abc from w_i. Thus $w_i = pabcqabcr$ for some $p, q, r \in A^*$, and either $w_{i-1} = pqabcr$ and $w_{i+1} = pabcqr$, or $w_{i-1} = pabcqr$ and $w_{i+1} = pqabcr$. Assume the

former case; the latter is similar. Then there is a sequence of elementary ρ-transitions

$$u = w_0 \leftrightarrow_\rho w_1 \leftrightarrow_\rho \ldots \leftrightarrow_\rho w_{i-1} = pqabcr \\ \leftrightarrow_\rho pqr \leftrightarrow pabcqr = w_{i+1} \leftrightarrow_\rho \ldots \leftrightarrow_\rho w_n = v.$$

Since $|pqr| < |w_i|$, this is a sequence where the number of intermediate words of maximum length is smaller, or (if no other intermediate word had length $|w_i|$) a smaller maximum length of intermediate words; in either case, this is a contradiction. Hence every element of M has a unique representative in N.

2.9 To define an assignment of generators $\varphi : A \to B_2$, proceed as follows. As noted in the question, $z\varphi$ must be the zero matrix $\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$. Furthermore, $(a\varphi)^2$ and $(b\varphi)^2$ must be the zero matrix. Calculating the squares of the available matrices shows that $a\varphi$ and $b\varphi$ must be in $\left\{\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}\right\}$. Since a and b can be swapped in $\sigma \cup \zeta$ to give the same set of defining relations, it does not matter which matrix we assign to each of $a\varphi$ and $b\varphi$. So define

$$a\varphi = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad b\varphi = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \quad z\varphi = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

Straightforward calculations show that B_2 satisfies all the defining relations in $\sigma \cup \zeta$ with respect to φ.

Let $N = \{z, a, b, ab, ba\}$. Let $w \in A^+$. If w contains a symbol z, then applying defining relations from ζ shows that (w, z) is a consequence of $\sigma \cup \zeta$. If w contains a^2 or b^2, then applying a single relation (a^2, z) or (b^2, z) introduces a symbol z, and so by the previous sentence (w, z) is a consequence of $\sigma \cup \zeta$. Finally, if w contains no a^2 or b^2 or z, then it consists of alternating symbols a and b, and so applying relations (aba, a) or (bab, b) transforms it to a word $u \in \{a, b, ab, ba\}$, and (w, u) is a consequence of $\sigma \cup \zeta$.

Lastly, $\varphi^+|_N$ is injective since the five words in $N = \{z, a, b, ab, ba\}$ correspond to the five matrices in B_2 (in the order listed in the question).

2.10 a) Suppose $c^\gamma b^\beta$ is idempotent. If $\gamma > \beta$, then

$$(c^\gamma b^\beta)^2 =_B c^\gamma b^\beta c^\gamma b^\beta =_B c^{\gamma+\gamma-\beta} b^\beta \neq_B c^\gamma b^\beta.$$

If $\gamma < \beta$, then

$$(c^\gamma b^\beta)^2 =_B c^\gamma b^\beta c^\gamma b^\beta =_B c^\gamma b^{\beta+\beta-\gamma} \neq_B c^\gamma b^\beta.$$

Hence $\gamma = \beta$. On the other hand, $(c^\gamma b^\gamma)^2 = c^\gamma b^\gamma c^\gamma b^\gamma =_B c^\gamma b^\gamma$ and so $c^\gamma b^\gamma$ is idempotent.

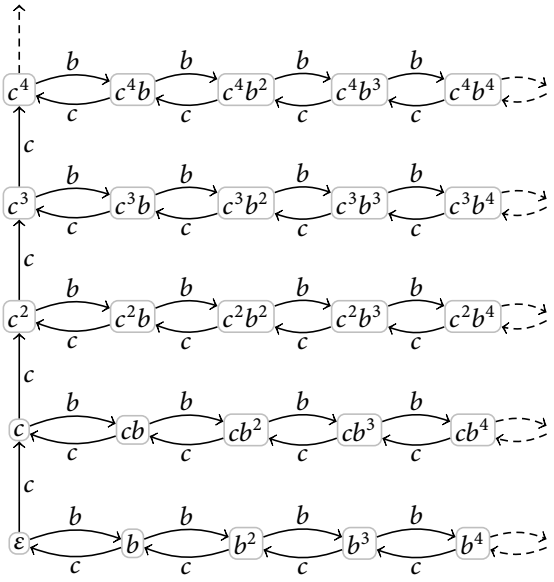


FIGURE S.4
Part of the Cayley graph of the bicyclic monoid.

b) Suppose first that c is right-invertible. Then there exists $c^\zeta b^\eta$ such that $cc^\zeta b^\eta =_B \varepsilon$. But this is impossible, since $c^{1+\zeta}b^\eta \neq_B \varepsilon$ since $1+\zeta > 0$. Now suppose that $c^\gamma b^\beta$, where $\gamma \geqslant 1$, has a right inverse x. Then $cc^{\gamma-1}b^\beta x =_B \varepsilon$ and so c is right-invertible, which is a contradiction. Hence if $c^\gamma b^\beta$ is right-invertible, then $\gamma = 0$. On the other hand, $b^\beta c^\beta =_B \varepsilon$ and so c^β is a right inverse for b^β.

2.11 The Cayley graph $\Gamma(B, \{b, c\})$ is shown in Figure S.4.

2.12 a) Suppose, with the aim of obtaining a contradiction, that $x^k = x^{k+m}$ for some $k, m \in \mathbb{N}$. Then $e = x^k y^k = x^{k+m}y^k = x^m$. Then $y = ey = x^m y = x^{m-1}e = x^{m-1}$ and so $yx = x^m = e$, which is a contradiction. So x is not periodic. Similarly y is not periodic.

b) Suppose $x^k = y^\ell$. Then $x^{k+\ell+1} = x^{\ell+1}y^\ell = x$. Since x is not periodic, this forces $k = \ell = 0$.

c) Suppose $y^k x^\ell = e$. Suppose, with the aim of obtaining a contradiction, that $\ell > 0$. Then $yx = eyx = y^k x^\ell yx = y^k x^{\ell-1}x = y^k x^\ell = e$, which is a contradiction. Thus $\ell = 0$, and so $y^{k+1} = ey = y$ and so $k = 0$ since y is not periodic.

d) Suppose, with the aim of obtaining a contradiction, that $y^k x^\ell = y^m x^n$ with either $k \neq m$ or $\ell \neq n$. Assume $k \neq m$; the other case is similar. Interchanging the two products if necessary, assume that $k < m$. Then $x^\ell = ex^\ell = x^k y^k x^\ell = x^k y^m x^n = ey^{m-k}x^n = y^{m-k}x^n$. If $\ell \geqslant n$, then $y^{m-k} = y^{m-k}x^n y^n = x^\ell y^n = x^{\ell-n}$, which contradicts part b). If $\ell \leqslant n$, then $e = x^\ell y^\ell = y^{m-k}x^n y^\ell = y^{m-k}x^{n-\ell}$, which contradicts part c).

e) Define $\varphi : B \to \langle x, y\rangle$ by $b\varphi = x$ and $y\varphi$. The given properties of e, x, and y show that φ is a well-defined homomorphism; it is clearly surjective; part d) shows that it is injective.

2.13 Let $e = \varepsilon\varphi$, $x = b\varphi$, and $y = c\varphi$. Since φ is a homomorphism, e, x, and y satisfy the conditions $ex = xe = x$, $ey = ye = y$, and $xy = e$. Note further that $S = \langle x, y\rangle$ since φ is surjective. If the condition $yx \neq e$ is also satisfied, then by Exercise 2.12, S is isomorphic to the bicyclic monoid. On the other hand, if $yx = e$, then every element of S is left- and right-invertible and so S is a group.

EXERCISES FOR CHAPTER 3

[See pages 68–70 for the exercises.]

3.1 Let G be a subgroup of a semigroup. Let $x, y \in G$. Let $p = x^{-1}y$ and $q = y^{-1}x$. Then $xp = y$ and $yq = x$. So $x \mathrel{\mathcal{R}} y$. Similarly $x \mathrel{\mathcal{L}} y$. Hence $x \mathrel{\mathcal{H}} y$.

3.2 Suppose $u, v \in A^*$ are such that $u \mathrel{\mathcal{R}} v$. Then there exist $p, q \in A^*$ such that $up = v$ and $vq = p$. Then $upq = p$, so $|u| + |p| + |q| = |u|$, and so $|p| = |q| = 0$. Thus $p = q = \varepsilon$ and so $u = v$. That is, $\mathcal{R}$ is the identity relation id_{A^*}. Similarly, the Green's relations $\mathcal{L}$, and $\mathcal{J}$ are the identity relation. Hence $\mathcal{H} = \mathcal{R} \sqcap \mathcal{L}$ and $\mathcal{D} = \mathcal{R} \sqcup \mathcal{L}$ are the identity relation.

3.3 a) Suppose $\sigma \mathrel{\mathcal{L}} \tau$. Then there exist $\pi, \rho \in \mathcal{T}_X$ such that $\pi\sigma = \tau$ and $\rho\tau = \sigma$. Therefore

$$\operatorname{im}\sigma = X\sigma \supseteq (X\pi)\sigma = \operatorname{im}(\pi\sigma) = \operatorname{im}\tau,$$

and similarly $\operatorname{im}\tau \supseteq \operatorname{im}(\rho\tau) = \operatorname{im}\sigma$. Hence $\operatorname{im}\sigma = \operatorname{im}\tau$.

Now suppose $\operatorname{im}\sigma = \operatorname{im}\tau$. For each $x \in X$, we have $x\tau \in \operatorname{im}\tau = \operatorname{im}\sigma$ and so we can define $x\pi$ to be some element of X such that $(x\pi)\sigma = x\tau$. Then $\pi\sigma = \tau$. Similarly we can define $\rho \in \mathcal{T}_X$ so that $\rho\tau = \sigma$. Hence $\sigma \mathrel{\mathcal{L}} \tau$.

b) Suppose $\sigma \mathrel{\mathcal{R}} \tau$. Then there exist $\pi, \rho \in \mathcal{T}_X$ such that $\sigma\pi = \tau$ and $\tau\rho = \sigma$. Therefore $(x, y) \in \ker\sigma \Rightarrow x\sigma = y\sigma \Rightarrow x\sigma\pi = y\sigma\pi \Rightarrow x\tau = y\tau \Rightarrow (x, y) \in \ker\tau$. Thus $\ker\sigma \subseteq \ker\tau$. Similarly, $\ker\tau \subseteq \ker\sigma$. Hence $\ker\sigma = \ker\tau$.

Now suppose $\ker\sigma = \ker\tau$. We aim to define $\pi \in \mathcal{T}_X$ such that $\sigma\pi = \tau$. For each $x \in \operatorname{im}\sigma$, choose $y_x \in X$ such that $y_x\sigma = x$. Note that each $z \in X$ is related by $\ker\sigma$ to $y_{z\sigma}$ and to no other y_x. Since $\ker\sigma = \ker\tau$, we have $(z, y_{z\sigma}) \in \ker\tau$ and so $z\tau = y_{z\sigma}\tau$. For each $x \in \operatorname{im}\sigma$, define $x\pi = y_x\tau$. For $x \notin \operatorname{im}\sigma$, let $x\pi$ be arbitrary. Then for all $z \in X$, we have $z\sigma \in \operatorname{im}\sigma$ and so $z\sigma\pi = y_{z\sigma}\tau = z\tau$; hence $\sigma\pi = \tau$. Similarly, we can define $\rho \in \mathcal{T}_X$ so that $\tau\rho = \sigma$. Hence $\sigma \mathrel{\mathcal{R}} \tau$.

c) Suppose $\sigma \mathrel{\mathcal{D}} \tau$. Then there exists $\upsilon \in \mathcal{T}_X$ such that $\sigma \mathrel{\mathcal{L}} \upsilon \mathrel{\mathcal{R}} \tau$. Since $\upsilon \mathrel{\mathcal{R}} \tau$, there exist $\pi, \rho \in \mathcal{T}_X$ such that $\upsilon\pi = \tau$ and $\tau\rho = \upsilon$.

Hence $\tau\rho\pi = \tau$ and $\upsilon\pi\rho = \upsilon$. Therefore $\rho|_{\operatorname{im}\tau} : \operatorname{im}\tau \to \operatorname{im}\upsilon$ and $\pi|_{\operatorname{im}\upsilon} : \operatorname{im}\upsilon \to \operatorname{im}\tau$ are mutually inverse bijections. So $|\operatorname{im}\upsilon| = |\operatorname{im}\tau|$. Since $\sigma \mathcal{L} \upsilon$, we have $\operatorname{im}\sigma = \operatorname{im}\upsilon$ and thus $|\operatorname{im}\sigma| = |\operatorname{im}\upsilon| = |\operatorname{im}\tau|$.

Now suppose $|\operatorname{im}\sigma| = |\operatorname{im}\tau|$. Then there is a bijection $\mu : \operatorname{im}\sigma \to \operatorname{im}\tau$. Extend μ to a map $\pi \in \mathcal{T}_X$ by defining $x\pi$ arbitrarily for $x \in X \setminus \operatorname{im}\sigma$. Similarly extend μ^{-1} to a map $\rho \in \mathcal{T}_X$. Let $\upsilon = \sigma\pi$. Then $\upsilon\rho = \sigma$, so $\upsilon \mathcal{R} \sigma$. Furthermore $\operatorname{im}\upsilon = \operatorname{im}(\sigma\pi) = \operatorname{im}(\sigma\mu) = \operatorname{im}\tau$, so $\upsilon \mathcal{L} \tau$ by part a). Hence $\sigma \mathcal{D} \tau$.

Suppose $\sigma \mathcal{J} \tau$. Then there exist $\pi, \rho, \pi', \rho' \in \mathcal{T}_X$ such that $\sigma = \pi\tau\rho$ and $\tau = \pi'\sigma\rho'$. Therefore

$$|\operatorname{im}\sigma| = |X\sigma| = |X\pi\tau\rho| \leqslant |X\tau\rho| \leqslant |X\tau| = |\operatorname{im}\tau|;$$

similarly $|\operatorname{im}\tau| \leqslant |\operatorname{im}\sigma|$. So $|\operatorname{im}\sigma| = |\operatorname{im}\tau|$. Hence $\sigma \mathcal{D} \tau$. Therefore $\mathcal{J} \subseteq \mathcal{D}$. Since $\mathcal{D} \subseteq \mathcal{J}$ in general, it follows that $\mathcal{D} = \mathcal{J}$.

3.4 Consider the following elements of $\mathcal{T}_{\{1,2,3\}}$:

$$\rho = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 2 \end{pmatrix}; \ \sigma = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 3 \end{pmatrix}; \ \tau = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 1 & 3 \end{pmatrix}.$$

Notice that

$$\rho\sigma = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 3 \end{pmatrix}; \ \rho\tau = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 1 & 1 \end{pmatrix}; \ \sigma\tau = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 3 \end{pmatrix}.$$

Thus $\operatorname{im}\sigma = \operatorname{im}\tau = \{1,3\}$, but $\operatorname{im}\rho\sigma = \{1,3\} \neq \{1\} = \operatorname{im}\rho\tau$ and so $(\sigma,\tau) \in \mathcal{L}$ but $(\rho\sigma,\rho\tau) \notin \mathcal{L}$ by Exercise 3.3(a). So $\mathcal{L}$ is not a left congruence in $\mathcal{T}_{\{1,2,3\}}$. Similarly, $\ker\rho = \ker\sigma$ but $\ker\rho\tau \neq \ker\sigma\tau$ and so $(\rho,\sigma) \in \mathcal{R}$ but $(\rho\tau,\sigma\tau) \notin \mathcal{R}$ by Exercise 3.3(b). So $\mathcal{R}$ is not a right congruence in $\mathcal{T}_{\{1,2,3\}}$.

3.5 Let $(\ell_1, r_1), (\ell_2, r_2) \in B$. Then

$$\begin{aligned}(\ell_1, r_1) \mathcal{R} (\ell_2, r_2) &\Rightarrow (\exists (k,s) \in B)\big((\ell_1, r_1)(k,s) = (\ell_2, r_2)\big) \\ &\Rightarrow (\exists (k,s) \in B)\big((\ell_1, s) = (\ell_2, r_2)\big) \\ &\Rightarrow \ell_1 = \ell_2.\end{aligned}$$

On the other hand, if $(\ell, r_1), (\ell, r_2) \in \{\ell\} \times R$, then $(\ell, r_1)(\ell, r_2) = (\ell, r_2)$ and $(\ell, r_2)(\ell, r_1) = (\ell, r_1)$ and so $(\ell, r_1) \mathcal{R} (\ell, r_2)$. So the $\mathcal{R}$-classes of B are the sets $\{\ell\} \times R$.

The result for $\mathcal{L}$-classes is proved similarly. Therefore

$$\begin{aligned}(\ell_1, r_1) \mathcal{H} (\ell_2, r_2) &\Leftrightarrow \big((\ell_1, r_1) \mathcal{L} (\ell_2, r_2)\big) \wedge \big((\ell_1, r_1) \mathcal{L} (\ell_2, r_2)\big) \\ &\Leftrightarrow (r_1 = r_2) \wedge (\ell_1 = \ell_2),\end{aligned}$$

and so $\mathcal{H}$ is the identity relation.

Finally, let $(\ell_1, r_1), (\ell_2, r_2) \in B$. Then $(\ell_1, r_1) \mathcal{R} (\ell_1, r_2) \mathcal{L} (\ell_2, r_2)$ and so $(\ell_1, r_1) \mathcal{D} (\ell_2, r_2)$. Hence B has consists of a single $\mathcal{D}$-class.

3.6 If $x \mathrel{\mathcal{R}} y$, then there exist $p, q \in S^1$ with $xp = y$ and $yq = x$. So $xpq = x$. Suppose that $pq \in S$. Then for any $z \in S$, we have $xpqz = xz$ and so $pqz = z$ by cancellativity. So pq is a left identity for S and in particular an idempotent. By Exercise 1.3, pq is an identity, which is a contradiction. So $pq \notin S$ and thus $pq = 1$, the adjoined identity of S^1. Hence $p = q = 1$ and so $x = y$. Thus $\mathcal{R} = \mathrm{id}_S$. Similarly $\mathcal{L} = \mathrm{id}_S$, and so $\mathcal{H} = \mathcal{R} \sqcap \mathcal{L} = \mathrm{id}_S$ and $\mathcal{D} = \mathcal{R} \sqcup \mathcal{L} = \mathrm{id}_S$.

3.7 Let $a, b, c, d, e, f \in \mathbb{R}$ with $a, b, c, d, e, f > 0$. Then

$$\begin{bmatrix} a & b \\ 0 & 1 \end{bmatrix} \begin{bmatrix} c & d \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} ac & ad+b \\ 0 & 1 \end{bmatrix};$$

since $ac > 0$ and $ad + b > 0$, we see that S is a subsemigroup of $M_2(\mathbb{R})$. Furthermore,

$$\det \begin{bmatrix} a & b \\ 0 & 1 \end{bmatrix} = a > 0;$$

thus every matrix in S is invertible; hence S is a subsemigroup of the general linear group $\mathrm{GL}_2(\mathbb{R})$ and so cancellative. Furthermore,

$$\begin{aligned} & \begin{bmatrix} e & f \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a & b \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} a & b \\ 0 & 1 \end{bmatrix} \\ \Rightarrow & \begin{bmatrix} ea & eb+f \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} a & b \\ 0 & 1 \end{bmatrix} \\ \Rightarrow & ea = a \wedge eb + f = b \\ \Rightarrow & e = 1 \wedge eb + f = b \\ \Rightarrow & e = 1 \wedge f = 0, \end{aligned}$$

which shows that S does not contain a left identity; thus S does not contain an identity. Finally, let $g, h \in \mathbb{R}$ with $g, h > 0$. Choose $f = 1$, $d = 0$, $c = h/(a+b)$, and $e = g/ca$. Then $c, d, e, f > 0$ and

$$\begin{aligned} & \begin{bmatrix} c & d \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a & b \\ 0 & 1 \end{bmatrix} \begin{bmatrix} e & f \\ 0 & 1 \end{bmatrix} \\ = & \begin{bmatrix} cae & caf + cb + d \\ 0 & 1 \end{bmatrix} \\ = & \begin{bmatrix} ca(g/ca) & (h/(a+b))a + (h/(a+b))b \\ 0 & 1 \end{bmatrix} \\ = & \begin{bmatrix} g & h \\ 0 & 1 \end{bmatrix}. \end{aligned}$$

Thus for any $x \in S$, we have $SxS = S$ and so S is simple. Hence $\mathcal{J} = S \times S$.

3.8 Suppose H_τ is a subgroup. Then $\tau^2 \in H_\tau$. In particular, $\tau \mathcal{D} \tau^2$ and so $|\operatorname{im}\tau| = |\operatorname{im}\tau^2|$.

Now suppose that $|\operatorname{im}\tau| = |\operatorname{im}(\tau^2)|$. First, notice that $\operatorname{im}(\tau^2) = X\tau^2 \subseteq X\tau = \operatorname{im}\tau$. Since $|\operatorname{im}(\tau^2)| = |\operatorname{im}\tau|$, we have $\operatorname{im}\tau^2 = \operatorname{im}\tau$ since $\operatorname{im}(\tau^2)$ and $\operatorname{im}\tau$ are finite (because X is finite). Also, $(x, y) \in \ker\tau \Rightarrow x\tau = y\tau \Rightarrow x\tau^2 = y\tau^2 \Rightarrow (x, y) \in \ker(\tau^2)$, and so $\ker\tau \subseteq \ker(\tau^2)$. So each $\ker(\tau^2)$-class is a union of $\ker\tau$-classes. Suppose, with the aim of obtaining a contradiction, that $\ker(\tau^2) - \ker\tau \neq \varnothing$. Then some $\ker(\tau^2)$-class is a union of at least two distinct $\ker\tau$-classes. So $\operatorname{im}(\tau^2) \subsetneq \operatorname{im}\tau$. Hence, since $\operatorname{im}(\tau^2)$ and $\operatorname{im}\tau$ are finite, $|\operatorname{im}(\tau^2)| < |\operatorname{im}\tau|$, which is a contradiction. So $\ker(\tau^2) = \ker\tau$. Hence $\tau \mathcal{L} \tau^2$ and $\tau \mathcal{R} \tau^2$ and so $\tau \mathcal{H} \tau^2$. Therefore H_τ is a subgroup.

3.9 First, let $x, y \in \{c^\gamma b^\beta : \beta \in \mathbb{N} \cup \{0\}\}$. Interchanging x and y if necessary, suppose $x = c^\gamma b^\beta$ and $y = c^\gamma b^\delta$ where $\beta \leqslant \delta$. Then $xb^{\delta-\beta} = y$ and $yc^{\delta-\beta} = x$. Hence $x \mathcal{R} y$.

Now suppose $c^\gamma b^\beta \mathcal{R} c^{\gamma+\eta} b^\delta$ for some $\eta > 0$. Then since $\mathcal{R}$ is a left congruence, we have $b^\beta =_B b^\gamma c^\gamma b^\beta \mathcal{R} b^\gamma c^{\gamma+\eta} b^\delta =_B c^\eta b^\delta$. Therefore there exists $p \in B$ such that $c^\eta b^\delta p =_B b^\beta$. Hence $c^\eta b^\delta p c^\beta =_B \varepsilon$ and so c^η is right-invertible, which contradicts Exercise 2.10(b). Hence $\{c^\gamma b^\beta : \beta \in \mathbb{N} \cup \{0\}\}$ is an $\mathcal{R}$-class.

Similarly, $\mathcal{L}$-classes are of the form $\{c^\gamma b^\beta : \gamma \in \mathbb{N} \cup \{0\}\}$. Finally, note that $c^\gamma b^\beta \mathcal{R} c^\gamma b^\delta \mathcal{L} c^\eta b^\delta$ and so $c^\gamma b^\beta \mathcal{D} c^\eta b^\delta$. Thus B consists of a single $\mathcal{D}$-class.

3.10 Let $e \in L \cap R$ be idempotent. Then e is a right identity for L and a left identity for R. For any $y \in R$, we have $ey = y$ and so $\rho_y|_L$ is a bijection from L to L_y. Let $z \in D$. Choose $y \in R \cap L_z$. Since $\rho_y|_L$ is a bijection, there exists $x \in L$ such that $z = x\rho_y|_L = xy \in LR$.

Hence $D \subseteq LR$. Let $x \in L$ and $y \in R$. Since $L \cap R$ contains the idempotent e, we have $xy \in L_y \cap R_x \subseteq D$ by Proposition 3.18. Hence $LR \subseteq D$.

3.11 Let $w \in \operatorname{Mon}\langle bc, c\rangle$. Then $w = \beta_1 \cdots \beta_n$, where each β_i is either bc or c. Let

$$\alpha_i = \begin{cases} a & \text{if } \beta_i = bc, \\ ab & \text{if } \beta_i = c. \end{cases}$$

Then

$$\begin{aligned} \alpha_n \cdots \alpha_1 w &= \alpha_n \cdots \alpha_1 \beta_1 \cdots \beta_n \\ &= \alpha_n \cdots \alpha_2 abc\beta_2 \cdots \beta_n \\ &=_M \alpha_n \cdots \alpha_2 \beta_2 \cdots \beta_n \\ &\;\;\vdots \\ &=_M \varepsilon. \end{aligned}$$

Clearly, $w\varepsilon =_M w$, so if $w \in \operatorname{Mon}\langle bc, c\rangle$, then $w \mathcal{L} \varepsilon$.

Now suppose $w \in N$ with $w \mathcal{L} \varepsilon$. Then there is a word $u \in N$ such that $uw =_M \varepsilon$. By Exercise 2.8, ε can be obtained from uw by deleting subwords abc. Neither u nor v contain subwords abc, so uw must have a subword abc across the 'boundary' of u and w. That is, we have either:

- $w = bcw'$ and $u = u'a$, with $u'w' =_M uw =_M \varepsilon$; or
- $w = cw'$ and $u = u'ab$, with $u'w' =_M uw =_M \varepsilon$.

Again, u' and w', being subwords of u and w, do not contain subwords abc, so the same reasoning applies. Proceeding by induction, we see that $w \in \mathrm{Mon}\langle bc, c\rangle$ (and $u \in \mathrm{Mon}\langle a, ab\rangle$, although that is not important). Hence if $w \mathcal{L} \varepsilon$, then $w \in \mathrm{Mon}\langle bc, c\rangle$.

Symmetrical reasoning shows that $w \mathcal{R} \varepsilon$ if and only if $w \in \mathrm{Mon}\langle a, ab\rangle$. Since $\mathcal{H} = \mathcal{L} \cap \mathcal{R}$, it follows that $w \mathcal{H} \varepsilon$ if and only if $w \in \mathrm{Mon}\langle bc, c\rangle \cap \mathrm{Mon}\langle a, ab\rangle = \{\varepsilon\}$.

Since ε is an idempotent, Exercise 3.10 shows that the $\mathcal{D}$-class of ε is $\mathrm{Mon}\langle bc, c\rangle \mathrm{Mon}\langle a, ab\rangle$.

If $w \in \mathrm{Mon}\langle bc, c\rangle \mathrm{Mon}\langle ab, a\rangle$, then $w \mathcal{D} \varepsilon$ and so $w \mathcal{J} \varepsilon$. If $w \in \mathrm{Mon}\langle bc, c\rangle b \mathrm{Mon}\langle ab, a\rangle$, then by the results for $\mathcal{L}$ and $\mathcal{R}$, there exist $p, q \in M$ such that $pwq =_M b$, so $apwqc =_M abc =_M \varepsilon$.

Now suppose $w \in N$ with $w \mathcal{J} \varepsilon$. Then there exist $u, v \in N$ such that $uwv =_M \varepsilon$. So ε can be obtained from uwv by deleting subwords abc. Any subwords abc in uwv must be across the boundaries of u and w and of v and w. Proceeding by induction as in the $\mathcal{L}$ case, we see that $w = \mathrm{Mon}\langle bc, c\rangle x \mathrm{Mon}\langle a, ab\rangle$, where x is either ε or a single letter b.

3.12 Since S is regular, $\mathcal{L}$-class and every $\mathcal{R}$-class of S contains an idempotent. Since there is only one idempotent in S, there is only one $\mathcal{R}$-class and only one $\mathcal{L}$-class in S. Hence $\mathcal{R} = \mathcal{L} = \mathcal{H} = S \times S$. So S consists of a single H-class, which contains an idempotent and is thus a subgroup.

3.13 a) Let $x \in R_1$. Then there exists $q \in M$ such that $xq = 1$. Since M is group-embeddable, $qx = 1$. Thus any element of R_1 is right- and left-invertible. On the other hand, if $x \in M$ is right-invertible, then $x \in R_1$. So $x \in R_1$ if and only if x is right- and left-invertible.

b) Suppose M has at least two $\mathcal{R}$-classes. Then $M \setminus R_1$ is non-empty. Let $x \in M \setminus H_1$. Then x is not right or left invertible. Suppose that $x^k \mathcal{R} x^\ell$ for some $k < \ell$. Then there exists $p \in M$ such that $x^k = x^\ell p$. Hence $x^{\ell-k} p = 1$ and so x has a right inverse $x^{\ell-k-1} p$. This is a contradiction. So all of the powers of x lie in different $\mathcal{R}$-classes.

EXERCISES FOR CHAPTER 4

[See pages 89–91 for the exercises.]

4.1 a) Define $\varphi : G \to \mathcal{M}[G; I, \Lambda; P]$ by $x \mapsto (1, xp_{11}^{-1}, 1)$. Then

$$\begin{aligned}(x\varphi)(y\varphi) &= (1, xp_{11}^{-1}, 1)(1, yp_{11}^{-1}, 1)\\ &= (1, xp_{11}^{-1}p_{11}yp_{11}^{-1}, 1)\\ &= (1, xyp_{11}^{-1}, 1)\\ &= (xy)\varphi.\end{aligned}$$

So φ is a homomorphism. Furthermore,

$$\begin{aligned}x\varphi = y\varphi &\Rightarrow (1, xp_{11}^{-1}, 1) = (1, yp_{11}^{-1}, 1)\\ &\Rightarrow xp_{11}^{-1} = yp_{11}^{-1}\\ &\Rightarrow x = y.\end{aligned}$$

So φ is injective. Finally, since G is a group, $(1, xp_{11}^{-1}, 1)$ will range over $\mathcal{M}[G; I, \Lambda, P] = \{1\} \times G \times \{1\}$ as x ranges over G. So φ is surjective. Hence φ is an isomorphism.

b) Let $M = \{e, z\}$ be a semilattice with $e > z$. Let $p_{\lambda i} = z$. Let (i, x, λ) and (i, y, λ) be arbitrary elements of $\mathcal{M}[M; I, \Lambda; P]$. Then

$$(i, x, \lambda)(i, y, \lambda) = (i, xp_{\lambda i}y, \lambda) = (i, xzy, \lambda) = (i, z, \lambda).$$

So $\mathcal{M}[M; I, \Lambda; P]$ is a null semigroup and so not isomorphic to M.

4.2 A completely simple semigroup is isomorphic to $\mathcal{M}[G; I, \Lambda; P]$ for some group G, index sets I and Λ, and matrix P. Suppose that we have $(i_1, g_1, \lambda_1)(i_2, g_2, \lambda_2) = (j_1, h_1, \mu_1)(j_2, h_2, \mu_2)$. The, by the definition of the product in $\mathcal{M}[G; I, \Lambda; P]$, we have $(i_1, g_1 p_{\lambda_1 i_2} g_2, \lambda_2) = (j_1, h_1 p_{\mu_1 j_2} h_2, \mu_2)$, and so

$$i_1 = j_1, \tag{S.13}$$
$$\lambda_2 = \mu_2, \tag{S.14}$$
$$g_1 p_{\lambda_1 i_2} g_2 = h_1 p_{\mu_1 j_2} h_2. \tag{S.15}$$

Let $q = (j_2, p_{\mu_1 j_2}^{-1} h_1^{-1} g_1, \lambda_1)$. Then

$$\begin{aligned}&(j_1, h_1, \mu_1)q\\ &= (j_1, h_1, \mu_1)(j_2, p_{\mu_1 j_2}^{-1} h_1^{-1} g_1, \lambda_1) && \text{[by definition of } q\text{]}\\ &= (j_1, h_1 p_{\mu_1 j_2} p_{\mu_1 j_2}^{-1} h_1^{-1} g_1, \lambda_1)\\ &= (j_1, g_1, \lambda_1)\\ &= (i_1, g_1, \lambda_1) && \text{[by (S.13)]}\end{aligned}$$

and

$$\begin{aligned}
&q(i_2, g_2, \lambda_2) \\
&= (j_2, p_{\mu_1 j_2}^{-1} h_1^{-1} g_1, \lambda_1)(i_2, g_2, \lambda_2) && \text{[by definition of } q\text{]} \\
&= (j_2, p_{\mu_1 j_2}^{-1} h_1^{-1} g_1 p_{\lambda_1 i_2} g_2, \lambda_2) \\
&= (j_2, p_{\mu_1 j_2}^{-1} h_1^{-1} h_1 p_{\mu_1 j_2} h_2, \mu_2) && \text{[by (S.14) and (S.15)]} \\
&= (j_2, h_2, \mu_2).
\end{aligned}$$

Hence $\mathcal{M}[G; I, \Lambda; P]$ is equidivisible.

4.3 a) Let $x, y \in S \simeq \mathcal{M}[G; I, \Lambda; P]$ with $x \mathcal{L} y$. Then by Proposition 4.12, $x = (i, g, \lambda)$ and $y = (j, h, \lambda)$ for some $i, j \in I$, $g, h \in G$, and $\lambda \in \Lambda$. Let $z = (k, f, \mu) \in S$. Then $zx = (k, f p_{\mu i} g, \lambda)$ and $zy = (k, f p_{\mu j} h, \lambda)$. Since $zx, zy \in I \times G \times \{\lambda\}$, we have $zx \mathcal{L} zy$. Hence $\mathcal{L}$ is left compatible. We already know $\mathcal{L}$ is a right congruence by Proposition 3.4(a). So $\mathcal{L}$ is a congruence. Similarly, $\mathcal{R}$ is a congruence and so $\mathcal{H} = \mathcal{L} \cap \mathcal{R}$ is a congruence.

b) Let $[(i, g, \lambda)]_{\mathcal{L}}, [(j, h, \mu)]_{\mathcal{L}} \in S/\mathcal{L}$. Then $[(i, g, \lambda)]_{\mathcal{L}}[(j, h, \mu)]_{\mathcal{L}} = [(i, g p_{\lambda j} h, \mu)]_{\mathcal{L}} = [(j, h, \mu)]_{\mathcal{L}}$ (since $(i, g p_{\lambda j} h, \mu)$ and (j, h, μ) are $\mathcal{L}$-related). Hence $S/\mathcal{L}$ is a right zero semigroup. Similarly $S/\mathcal{R}$ is a left zero semigroup.

c) Define a map $\varphi : S/\mathcal{H} \to S/\mathcal{R} \times S/\mathcal{L}$ by

$$[(i, g, \lambda)]_{\mathcal{H}}\varphi = \big([(i, g, \lambda)]_{\mathcal{R}}, [(i, g, \lambda)]_{\mathcal{L}}\big).$$

Using the fact that $\mathcal{H} = \mathcal{L} \cap \mathcal{R}$, it is easy to show that this map is well-defined and injective. It is clearly surjective, and is a homomorphism since $\mathcal{R}$ and $\mathcal{L}$ are congruences. So $S/\mathcal{H} \simeq S/\mathcal{R} \times S/\mathcal{L}$.

4.4 Since S is completely simple, $S \simeq \mathcal{M}[G; I, \Lambda, P]$. Hence $|S| = |I| \times |G| \times |\Lambda|$.

a) Since $p = |I| \times |G| \times |\Lambda|$, one of the following three cases must hold:

i) $|I| = p$, $|G| = 1$, and $|\Lambda| = 1$. Since G is trivial and $|\Lambda| = 1$, the $\mathcal{R}$-classes of S are single elements by Proposition 4.12(c). Thus $S \simeq S/\mathcal{R}$ is a left zero semigroup by Exercise 4.3(b).

ii) $|I| = 1$, $|G| = 1$, and $|\Lambda| = p$. This is similar to case i), and shows that S is a right zero semigroup.

iii) $|I| = 1$, $|G| = p$, and $|\Lambda| = 1$. Then S is a group by Exercise 4.1.

b) Since $pq = |I| \times |G| \times |\Lambda|$, one of the following cases must hold (interchanging p and q if necessary):

i) $|I| = pq$, $|G| = 1$, and $|\Lambda| = 1$. As in part a)i), $S \simeq S/\mathcal{R}$ is a left zero semigroup and so a left group by Theorem 4.19. [We could also use the fact that S has only one $\mathcal{L}$-class and apply Theorem 4.19.]

ii) $|I| = p$, $|G| = q$, and $|\Lambda| = 1$. Then $S = I \times G \times \{\lambda\}$. Thus S has only one $\mathcal{L}$-class and so is a left group by Theorem 4.19.

iii) $|I| = p$, $|G| = 1$, and $|\Lambda| = q$. Then the $\mathcal{H}$-classes of S are single elements by Proposition 4.12(d). So $S \simeq S/\mathcal{H}$ is a rectangular band by Exercise 4.3(c)

iv) $|I| = 1$, $|G| = pq$, and $|\Lambda| = 1$. As in part a)iii), S is a group (and thus both a left and a right group).

v) $|I| = 1$, $|G| = p$, and $|\Lambda| = q$. This is dual to case ii), and shows that S is a right group.

vi) $|I| = 1$, $|G| = 1$, and $|\Lambda| = pq$. This is dual to case i), $S \simeq S/\mathcal{L}$ is a right zero semigroup and so a right group.

4.5 a) Let $z \in S$. Then $zz^{-1}\ \mathcal{R}\ z$ and $z^{-1}z\ \mathcal{L}\ z$. So $zz^{-1} = z^{-1}z\ \mathcal{H}\ z$. Similarly $zz^{-1} = z^{-1}z\ \mathcal{H}\ z^{-1}$. So $z\ \mathcal{H}\ z^{-1}$. Since every $\mathcal{H}$-class of S is a subgroup, z^{-1} is the unique group inverse of z in this subgroup. The $\mathcal{H}$-class of $z\varphi$ is also a subgroup and $(z\varphi)^{-1}$ is the unique group inverse of $z\varphi$ in this subgroup. Then $\varphi|_{H_z}$ is a group homomorphism into the subgroup $H_{z\varphi}$ and so $z^{-1}\varphi = (z\varphi)^{-1}$.

b) There are many possible examples. Let $S = \{s_1, s_2\}$ and $T = \{t_1, t_2\}$ be left zero semigroups. Define $^{-1}$ on S by $s_1^{-1} = s_2$ and $s_2^{-1} = s_1$. Define $^{-1}$ on T by $t_1^{-1} = t_1$ and $t_2^{-1} = t_2$. In both cases, $^{-1}$ satisfies $(x^{-1})^{-1} = x$ and $xx^{-1}x = x$. Define $\varphi : S \to T$ by $s_1\varphi = t_1$ and $s_2\varphi = t_2$. Then $(s_1\varphi)^{-1} = t_1^{-1} = t_1$ but $s_1^{-1}\varphi = s_2\varphi = t_2$.

4.6 a) i) The isomorphism φ maps non-zero $\mathcal{R}$-classes bijectively to non-zero $\mathcal{R}$-classes. Since the $\mathcal{R}$-classes of $\mathcal{M}_0[G; I, \Lambda; P]$ are sets of the form $\{i\}\times G\times\Lambda$ and the $\mathcal{R}$-classes of $\mathcal{M}_0[H; J, M; Q]$ are sets of the form $\{j\} \times G \times M$, there must be a bijection $\alpha : I \to J$ such that $(i, a, \lambda)\varphi \in \{i\alpha\} \times H \times M$. Similarly there is a bijection $\beta : \Lambda \to M$ such that $(i, a, \lambda)\varphi \in I \times H \times \{\lambda\beta\}$. Combining these statements shows that $(i, a, \lambda)\varphi \in \{i\alpha\} \times H \times \{\lambda\beta\}$. Since φ must map group $\mathcal{H}$-classes to group $\mathcal{H}$-classes, we have $p_{\lambda i} \neq 0$ if and only if $p_{(\lambda\beta)(i\alpha)} \neq 0$.

ii) Let $\gamma : G \to \{1\}\times G\times\{1\}$ be defined by $x\gamma = (1, p_{11}^{-1}x, 1)$. Then $(x\gamma)(y\gamma) = (1, p_{11}^{-1}x, 1)(1, p_{11}^{-1}y, 1) = (1, p_{11}^{-1}xp_{11}p_{11}^{-1}y, 1) = (1, p_{11}^{-1}xy, 1) = (xy)\gamma$, so γ is a homomorphism. Furthermore, γ is injective since $x\gamma = y\gamma \Rightarrow (1, p_{11}^{-1}x, 1) = (1, p_{11}^{-1}y, 1) \Rightarrow p_{11}^{-1}x = p_{11}^{-1}y \Rightarrow x = y$. Finally, γ is surjective since for any $(1, x, 1) \in \{1\} \times G \times \{1\}$, we have $(p_{11}x)\gamma = (1, x, 1)$. So γ is an isomorphism.

Similarly, the map $\eta : H \to \{1\alpha\} \times H \times \{1\beta\}$ defined by $x\eta = (1\alpha, q_{(1\beta)(1\alpha)}^{-1}x, 1\beta)$ is an isomorphism.

By part i), $\varphi|_{\{1\}\times G\times\{1\}} : \{1\} \times G \times \{1\} \to \{1\alpha\} \times H \times \{1\beta\}$ is an isomorphism, so the composition $\vartheta = \gamma\varphi\eta^{-1} = \gamma\varphi|_{\{1\}\times G\times\{1\}}\eta^{-1}$ is an isomorphism from G to H.

iii) First,

$$\begin{aligned}(i, 1_G, 1)(1, p_{11}^{-1}x, 1)(1, p_{11}^{-1}, 1) &= (i, 1_G p_{11} p_{11}^{-1} x p_{11} p_{11}^{-1}) \\ &= (i, x, \lambda).\end{aligned}$$

Now, for all $x \in G$,

$$(1, p_{11}^{-1}x, 1)\varphi = x\gamma\varphi = x\vartheta\eta = (1\alpha, q_{(1\beta)(1\alpha)}^{-1}(x\vartheta), 1\beta).$$

Therefore for any $x \in G$,

$$\begin{aligned}&(i, x, \lambda)\varphi \\ &= \big((i, 1_G, 1)(1, p_{11}^{-1}x, 1)(1, p_{11}^{-1}, \lambda)\big)\varphi \\ &= (i, 1_G, 1)\varphi(1, p_{11}^{-1}x, 1)\varphi(1, p_{11}^{-1}, \lambda)\varphi \\ &= (i\alpha, u_i, 1\beta)(1\alpha, q_{(1\beta)(1\alpha)}^{-1}(x\vartheta), 1\beta)(1\alpha, q_{(1\beta)(1\alpha)}^{-1}v_\lambda, \lambda\beta) \\ &= (i\alpha, u_i q_{(1\beta)(1\alpha)} q_{(1\beta)(1\alpha)}^{-1}(x\vartheta) q_{(1\beta)(1\alpha)} q_{(1\beta)(1\alpha)}^{-1} v_\lambda, \lambda\beta) \\ &= (i\alpha, u_i(x\vartheta)v_\lambda, \lambda\beta)\varphi.\end{aligned}$$

Hence

$$\begin{aligned}&(i\alpha, u_i(p_{\lambda i}\vartheta)v_\lambda, \lambda\beta) \\ &= (i, p_{\lambda i}, \lambda)\varphi \\ &= \big((i, 1_G, \lambda)(i, 1_G, \lambda)\big)\varphi \\ &= (i, 1_G, \lambda)\varphi(i, 1_G, \lambda)\varphi \\ &= (i\alpha, u_i v_\lambda, \lambda\beta)(i\alpha, u_i v_\lambda, \lambda\beta) \\ &= (i\alpha, u_i v_\lambda q_{(\lambda\beta)(i\alpha)} u_i v_\lambda, \lambda\beta);\end{aligned}$$

thus $p_{\lambda i}\vartheta = v_\lambda q_{(\lambda\beta)(i\alpha)} u_i$ by cancellativity in H.

b) Define a map $\varphi : \mathcal{M}_0[G; I, \Lambda; P] \to \mathcal{M}_0[H; J, M; Q]$ by

$$(i, x, \lambda)\varphi = (i\alpha, u_i(x\vartheta)v_\lambda, \lambda\beta), \quad \text{and} \quad 0\varphi = 0.$$

Then φ is a homomorphism since

$$\begin{aligned}&(i, x, \lambda)\varphi(i', y, \lambda')\varphi \\ &= (i\alpha, u_i(x\vartheta)v_\lambda, \lambda\beta)(i'\alpha, u_{i'}(y\vartheta)v_{\lambda'}, \lambda') \\ &= (i\alpha, u_i(x\vartheta)v_\lambda q_{(\lambda\beta)(i'\alpha)} u_{i'}(y\vartheta)v_{\lambda'}, \lambda') \\ &= (i\alpha, u_i(x\vartheta)(p_{\lambda i'}\vartheta)(y\vartheta)v_{\lambda'}, \lambda') \\ &= (i\alpha, u_i((x p_{\lambda i'} y)\vartheta)v_{\lambda'}, \lambda') \\ &= (i, x p_{\lambda i'} y, \lambda')\varphi \\ &= \big((i, x, \lambda)(i', y, \lambda')\big)\varphi.\end{aligned}$$

Furthermore, φ is a bijection since α, β, and ϑ are all bijections. So φ is an isomorphism from $\mathcal{M}_0[G; I, \Lambda; P]$ to $\mathcal{M}_0[H; J, M; Q]$.

4.7 Suppose P is regular. Then $S = \mathcal{M}_0[G; I, \Lambda; P]$ is completely simple and so regular by the Lemma 4.6(b). [Alternatively: Since P contains some non-zero element $p_{\lambda i}$, the element $(i, p_{\lambda i}^{-1}, \lambda)$ is idempotent and thus regular. Thus the $\mathcal{D}$-class $I \times G \times \Lambda$ is regular by Proposition 3.19.]

Suppose P is not regular. Then P has a row or a column all of whose entries are 0. Suppose all the entries in the row indexed by λ are 0; the reasoning for columns is similar. Let $(i, x, \lambda) \in I \times G \times \{\lambda\}$. Then for $(j, y, \mu) \in \mathcal{M}_0[G; I, \Lambda; P] \setminus \{0\}$, we have $(i, x, \lambda)(j, y, \mu) = 0$ since $p_{\lambda j} = 0$. Hence there is no element $z \in \mathcal{M}_0[G; I, \Lambda; P]$ with $(i, x, \lambda)z(i, x, \lambda) = (i, x, \lambda)$. Thus S is not regular.

4.8 a) Since S satisfies $\min_{\mathcal{L}}$, the set of $\mathcal{L}$-classes that are not equal to $\{0\}$ there is a minimal element. Let L_x be such a minimal $\mathcal{L}$-class not equal to $\{0\}$. Then Sx is a left ideal not equal to $\{0\}$. Suppose L is some left ideal contained in Sx and not equal to $\{0\}$. Pick $y \in L \setminus \{0\}$. Then $Sy \subseteq Sx$ and so $L_y \subseteq L_x$. Since L_x is minimal among non-$\{0\}$ $\mathcal{L}$-classes, $L_x = L_y$ and so $Sx = Sy$. So Sx must be a 0-minimal left ideal.

b) i) Let $x \in K \setminus \{0\}$. Then Sx is a left ideal of S and is contained in K. Since K is 0-minimal, either $Sx = K$ or $Sx = \{0\}$. Suppose, with the aim of obtaining a contradiction, that $Sx = \{0\}$. Then $\{0, x\}$ is a left ideal of S contained in K and not equal to $\{0\}$. Since K is 0-minimal, $K = \{x, 0\}$. But then $K^2 = \{0\}$, which is a contradiction. So $K = Sx$.

ii) It is immediate that Lx is a left ideal. Suppose $K \neq \{0\}$ is a left ideal contained in Lx. Let $J = \{\, y \in L : yx \in K \,\}$. Then $J \subseteq L$ and J is a left ideal, since

$$\begin{aligned} & y \in J \wedge s \in S \\ \Rightarrow\ & y \in L \wedge yx \in K \wedge s \in S && \text{[by definition of } J\text{]} \\ \Rightarrow\ & sy \in L \wedge syx \in K && \text{[since } L \text{ and } K \text{ are left ideals]} \\ \Rightarrow\ & sy \in J. && \text{[by definition of } J\text{]} \end{aligned}$$

Since L is 0-minimal and $J \neq \{0\}$, we have $J = L$ and so $Jx = Lx$. Furthermore, $Jx \subseteq K$ by the definition of J and $K \subseteq Lx$ by the definition of K, and so $K = Jx = Lx$. Hence Lx is 0-minimal.

iii) Note that LS is an ideal since $SLSS \subseteq (SL)(S^2) \subseteq LS$. So, since S is 0-simple, either $LS = \{0\}$ or $LS = S$. Suppose, with the aim of obtaining a contradiction, that $LS = \{0\}$. Then $LS \subseteq L$ and so L is an ideal. Since $L \neq \{0\}$, we have $L = S$. Hence $S^2 = LS = \{0\}$ and so S is null, which contradicts S being 0-simple. Therefore $LS = S$. So there exists $x \in S$ with $Lx \neq \{0\}$.

iv) The set M is a union of 0-minimal left ideals and is thus itself a left ideal. By part iii), $M \neq \{0\}$. Let $m \in M$ and $t \in S$. Then

$m \in Lx$ for some $x \in S$ and so $mt \in Lxt \subseteq M$. Hence $MS \subseteq M$ and so M is also a right ideal. So M is an ideal and not equal to $\{0\}$. Since S is 0-simple, we have $M = S$.

v) Let L be a 0-minimal left ideal. For any 0-minimal right ideal R, the set LR is an ideal and hence, since S is 0-simple, either $LR = \{0\}$ or $LR = S$. By part iii), there exists some $x \in S$ with $Lx \neq \{0\}$. By the dual version of part iv), x lies in some 0-minimal right ideal. Fix a 0-minimal right ideal R containing x. Then $LR \neq \{0\}$ and so $LR = S$.

Notice that since R is a right ideal, $RL \subseteq R$. Similarly, $RL \subseteq L$. Let $x \in RL \setminus \{0\} \subseteq R \setminus \{0\}$. Then $R = xS$ by the dual version of part i). Since $S = LR = LxS$, we have $Lx \neq \{0\}$ and so Lx is a 0-minimal left ideal by part ii). However, $Lx \subseteq L$ since $x \in RL \subseteq L$. Therefore, since L is 0-minimal and $Lx \neq \{0\}$, we have $Lx = L$ and so $RLx = RL$. Similarly $xRL = RL$. Hence RL is a group with a zero adjoined by Exercise 1.21.

vi) Let f be a non-zero idempotent in S with $f \preccurlyeq e$. Then $ef = fe = f$. Since $e \in RL \subseteq R \cap L$, it follows from part i) and its dual version that $R = eS$ and $L = Se$. Hence $f = efe \in eSe = eS^2e = (eS)(Se) = RL$, since $S^2 = S$ by Lemma 3.6. Since $RL \setminus \{0\}$ is a group, $e = f$. So e is a primitive idempotent. Hence S is completely 0-simple.

4.9 Let R be a right ideal of S. Let $r \in R$ and $\ell \in G$. Then $r\ell \in R \cap G$ since R is a right ideal and G is a left ideal. So $R \cap G \neq \varnothing$. Then $R \cap G$ is a right ideal of G, since R is a right ideal and G is a subgroup. But G is a group, and thus its only right ideal is G itself. Hence $R \cap G = G$, and so $G \subseteq R$. In particular, $1_G \in R$. Let $x \in S$. Then $1_G x = 1_G 1_G x$ since 1_G is idempotent, and so $x = 1_G x$ since S is left-cancellative. Therefore $x = 1_G x \in 1_G S \subseteq RS \subseteq R$. Hence $S \subseteq R$ and so $S = R$. Therefore S does not contain any proper right ideals and so is right simple. Since it is also left-cancellative, S is a right group.

EXERCISES FOR CHAPTER 5

[See pages 116–119 for the exercises.]

5.1 Let $\tau = \begin{pmatrix} 1 & 2 \\ 2 & * \end{pmatrix}$ and $\zeta = \begin{pmatrix} 1 & 2 \\ * & * \end{pmatrix}$. Then $\tau\tau = \tau\zeta = \zeta\tau = \zeta\zeta = \zeta$. So $T = \{\tau, \zeta\}$ is a null semigroup and τ does not have an inverse in T. [Of course, τ does have an inverse in $\mathcal{I}_X$; indeed $\tau^{-1} = \begin{pmatrix} 1 & 2 \\ * & 1 \end{pmatrix}$.]

5.2 Let $\sigma_1, \sigma_2 \in S$. Then there exist subgroups H_1, H_1', H_2, and H_2' of G such that $\sigma_1 : H_1 \to H_1'$ and $\sigma_2 : H_2 \to H_2'$ are isomorphisms.

Then $\operatorname{dom}(\sigma_1\sigma_2) = (\operatorname{im}\sigma_1 \cap \operatorname{dom}\sigma_2)\sigma_1^{-1} = (H_1' \cap H_2)\sigma_1^{-1}$. Now, $H_1' \cap H_2$ is a subgroup of G. (In particular, it contains 1_G and so is non-empty.) Thus $\operatorname{dom}(\sigma_1\sigma_2)$ is a subgroup of G and so $\operatorname{im}(\sigma_1\sigma_2)$ is also a subgroup of G. So $\sigma_1\sigma_2 \in S$. Thus S is a subsemigroup of $\mathcal{I}_G$. Furthermore, $\sigma_1^{-1} : H_1' \to H_1$ is also an isomorphism; thus $\sigma_1^{-1} \in S$. Thus S is an inverse subsemigroup of $\mathcal{I}_G$.

5.3 a) Suppose that $\sigma \mathcal{L} \tau$. Then there exist $\pi, \rho \in \mathcal{I}_X$ such that $\pi\sigma = \tau$ and $\rho\tau = \sigma$. Therefore

$$\operatorname{im}\sigma = X\sigma \supseteq (X\pi)\sigma = \operatorname{im}(\pi\sigma) = \operatorname{im}\tau,$$

and similarly $\operatorname{im}\tau \supseteq \operatorname{im}(\rho\tau) = \operatorname{im}\sigma$. Hence $\operatorname{im}\sigma = \operatorname{im}\tau$.

Now suppose that $\operatorname{im}\sigma = \operatorname{im}\tau$. Let $\pi = \tau\sigma^{-1}$. Then

$$\pi\sigma = \tau\sigma^{-1}\sigma = \tau\mathrm{id}_{\operatorname{im}\sigma} = \tau\mathrm{id}_{\operatorname{im}\tau} = \tau.$$

Similarly, let $\rho = \sigma\tau^{-1}$; then $\rho\tau = \sigma$. Hence $\sigma \mathcal{L} \tau$.

b) Suppose that $\sigma \mathcal{R} \tau$. Then there exist $\pi, \rho \in \mathcal{I}_X$ such that $\sigma\pi = \tau$ and $\tau\rho = \sigma$. Therefore

$$\begin{aligned}\operatorname{dom}\tau = \operatorname{dom}\sigma\pi &= (\operatorname{im}\sigma \cap \operatorname{dom}\pi)\sigma^{-1}\\ &\subseteq (\operatorname{im}\sigma)\sigma^{-1} = \operatorname{dom}\sigma\end{aligned}$$

and similarly $\operatorname{dom}\sigma \subseteq (\operatorname{im}\tau)\tau^{-1} = \operatorname{dom}\tau$. Thus $\operatorname{dom}\sigma = \operatorname{dom}\tau$.

Now suppose that $\operatorname{dom}\sigma = \operatorname{dom}\tau$. Let $\pi = \sigma^{-1}\tau$. Then

$$\sigma\pi = \sigma\sigma^{-1}\tau = \mathrm{id}_{\operatorname{dom}\sigma}\tau = \mathrm{id}_{\operatorname{dom}\tau}\tau = \tau.$$

Similarly, let $\rho = \tau^{-1}\sigma$; then $\tau\rho = \sigma$. Hence $\sigma \mathcal{R} \tau$.

c) Suppose that $\sigma \mathcal{D} \tau$. Then there exists $\upsilon \in \mathcal{I}_X$ such that $\sigma \mathcal{L} \upsilon \mathcal{R} \tau$, and so,

$$\begin{aligned}|\operatorname{dom}\sigma| &= |\operatorname{im}\sigma| && \text{[since } \sigma \text{ is a partial bijection]}\\ &= |\operatorname{im}\upsilon| && \text{[by part a)]}\\ &= |\operatorname{dom}\upsilon| && \text{[since } \upsilon \text{ is a partial bijection]}\\ &= |\operatorname{dom}\tau|. && \text{[by part b)]}\end{aligned}$$

Now suppose that $|\operatorname{dom}\sigma| = |\operatorname{dom}\tau|$. Then there is a bijection $\pi : \operatorname{dom}\sigma \to \operatorname{dom}\tau$. Note that $\pi \in \mathcal{I}_X$. Let $\upsilon = \pi^{-1}\sigma$. Then $\sigma = \pi\upsilon$, and so $\sigma \mathcal{L} \upsilon$. Furthermore,

$$\begin{aligned}\operatorname{dom}\upsilon &= \operatorname{dom}(\pi^{-1}\sigma)\\ &= (\operatorname{im}\pi^{-1} \cap \operatorname{dom}\sigma)\pi\\ &= (\operatorname{dom}\pi \cap \operatorname{dom}\sigma)\pi\\ &= (\operatorname{dom}\sigma)\pi\\ &= \operatorname{dom}\tau,\end{aligned}$$

and so $\upsilon \mathcal{R} \tau$ by part b). Hence $\sigma \mathcal{D} \tau$.

Suppose $\sigma \mathcal{J} \tau$. Then there exist $\pi, \rho, \pi', \rho' \in \mathcal{I}_X$ such that $\sigma = \pi\tau\rho$ and $\tau = \pi'\sigma\rho'$. Therefore

$$\begin{aligned} |\operatorname{dom}\sigma| = |\operatorname{im}\sigma| = |X\sigma| &= |X\pi\tau\rho| \\ &\leqslant |X\tau\rho| = |X\tau| = |\operatorname{im}\tau| = |\operatorname{dom}\tau|; \end{aligned}$$

similarly, $|\operatorname{dom}\tau| \leqslant |\operatorname{dom}\sigma|$. Thus $|\operatorname{dom}\sigma| = |\operatorname{dom}\tau|$. Hence $\sigma \mathcal{D} \tau$. Therefore $\mathcal{J} \subseteq \mathcal{D}$ and so $\mathcal{D} = \mathcal{J}$.

5.4 a) Since $\operatorname{im}\pi = \operatorname{dom}\beta$, it follows that

$$\begin{aligned} \operatorname{dom}(\pi\beta) &= (\operatorname{im}\pi \cap \operatorname{dom}\beta)\pi^{-1} \\ &= (\operatorname{im}\pi)\pi^{-1} \\ &= \operatorname{dom}\pi \\ &= \operatorname{dom}\gamma. \end{aligned}$$

Hence $\pi\beta \mathcal{R} \gamma$ by Exercise 5.3(b). Thus there exists $\rho' \in \mathcal{I}_X$ such that $\pi\beta\rho' = \gamma$. Since $|\operatorname{im}\beta| = |\operatorname{dom}\beta| = n-1 = |\operatorname{dom}\gamma| = |\operatorname{im}\gamma|$, it follows that $\operatorname{dom}\rho' \supseteq \operatorname{im}\beta$. Extend ρ' to a permutation $\rho \in \mathcal{S}_X$. Then ρ and ρ' agree on $\operatorname{im}\beta$. Hence $\pi\beta\rho = \pi\beta\rho' = \gamma$.

Since $\pi, \rho \in \mathcal{S}_X = \langle \tau, \zeta \rangle$, it follows from the previous paragraph that $J_{n-1} \subseteq \mathcal{S}_X\beta\mathcal{S}_X \subseteq \langle \tau, \zeta, \beta \rangle$.

b) Let $\sigma \in J_k$. Pick $x \in X \setminus \operatorname{dom}\sigma$ and $y \in X \setminus \operatorname{im}\sigma$ and extend σ to σ' by defining $x\sigma' = y$. Then $\sigma' \in J_{k+1}$, and $\sigma = \sigma'\mathrm{id}_{X-\{x\}} \in J_{k+1}J_{n-1}$. Hence $J_k \subseteq J_{k+1}J_{n-1}$.

By induction on k, we see that $J_k \subseteq J_{n-1}^{n-k} \subseteq \langle \tau, \zeta, \beta \rangle$. Since this holds for $k = 0, \ldots, n-1$, and since obviously $J_n = \mathcal{S}_X = \langle \tau, \zeta \rangle \subseteq \langle \tau, \zeta, \beta \rangle$, it follows that $\mathcal{I}_X = \bigcup_{k=0}^{n} J_k \subseteq \langle \tau, \zeta, \beta \rangle$.

5.5 a) Let $M = \langle \tau, \tau^{-1} \rangle$. Note that $\tau\tau^{-1} = \mathrm{id}_X$, so M is a monoid. We will use Method 2.9 to prove that M is defined by $\operatorname{Mon}\langle b, c \mid (bc, \varepsilon) \rangle$.

Define $\varphi : \{b, c\} \to M$ by $b\varphi = \tau$ and $c\varphi = \tau^{-1}$. Then M satisfies the defining relation with respect to φ since $(bc)\varphi^* = \tau\tau^{-1} = \mathrm{id}_X = \varepsilon\varphi^*$. Let $N = \{ c^i b^j : i \in \mathbb{N} \cup \{0\} \}$; any word in $\{b, c\}^*$ can be transformed to one in N by applying the defining relation to delete subwords bc. Finally, let $x \in X \setminus X\tau$ (note that such an x exists since $\operatorname{im}\tau \subsetneq X$). Then for $k \in \mathbb{N} \cup \{0\}$, we have $x\tau^k \in X\tau^k \setminus X\tau^{k+1}$. In particular, the $x\tau^k$ are all distinct. Note that $x \notin \operatorname{im}\tau = \operatorname{dom}\tau^{-1}$. Thus $(x\tau^k)(c^i b^j)\varphi^* = x\tau^k\tau^{-i}\tau^j$ is defined if and only if $k \geqslant i$, in which case it is equal to $x\tau^{k-i+j}$. So the minimum k for which $(x\tau^k)(c^i b^j)\varphi^*$ is defined is i, and the image of $x\tau^i$ under $(c^i b^j)\varphi^*$ is $x\tau^j$. So $(c^i b^j)\varphi^*$ determines i and j, and so $\varphi^*|_N$ is injective. This completes the proof.

[This proof is essentially just Example 2.11(b) rephrased in terms of partial bijections.]

b) Let $M = \langle \{ \tau_i, \tau_i^{-1} : i \in I \} \rangle$. Note that $\tau_i \tau_i^{-1} = \mathrm{id}_X$, so M is a monoid. For any $i_1, \dots, i_k \in I$ and $\epsilon_1, \dots, \epsilon_k \in \{1, -1\}$,

$$\begin{aligned}
&\tau_{i_1}^{\epsilon_1} \cdots \tau_{i_k}^{\epsilon_k} \tau_{i_k}^{-\epsilon_k} \cdots \tau_{i_1}^{-\epsilon_1} \tau_{i_1}^{\epsilon_1} \cdots \tau_{i_k}^{\epsilon_k} \\
&= \tau_{i_1}^{\epsilon_1} \cdots \tau_{i_{k-1}}^{\epsilon_{k-1}} \mathrm{id}_X \tau_{i_{k-1}}^{-\epsilon_{k-1}} \cdots \tau_{i_1}^{-\epsilon_1} \tau_{i_1}^{\epsilon_1} \cdots \tau_{i_k}^{\epsilon_k} \\
&= \tau_{i_1}^{\epsilon_1} \cdots \tau_{i_{k-1}}^{\epsilon_{k-1}} \tau_{i_{k-1}}^{-\epsilon_{k-1}} \cdots \tau_{i_1}^{-\epsilon_1} \tau_{i_1}^{\epsilon_1} \cdots \tau_{i_k}^{\epsilon_k} \\
&\ \ \vdots = \qquad\qquad\qquad\qquad\qquad\qquad \mathrm{id}_X \tau_{i_1}^{\epsilon_1} \cdots \tau_{i_k}^{\epsilon_k} \\
&= \tau_{i_1}^{\epsilon_1} \cdots \tau_{i_k}^{\epsilon_k}.
\end{aligned}$$

So $(\tau_{i_1}^{\epsilon_1} \cdots \tau_{i_k}^{\epsilon_k})^{-1} = \tau_{i_k}^{-\epsilon_k} \cdots \tau_{i_1}^{-\epsilon_1} \in M$. Hence M is an inverse monoid. Furthermore, for $i, j \in I$ with $i \neq j$ since $\operatorname{im} \tau_i$ and $\operatorname{im} \tau_j = \operatorname{dom} \tau_j^{-1}$ are disjoint, $\tau_i \tau_j^{-1} = \varnothing$, and $\varnothing$ is a zero for $\mathcal{B}_X$ and thus for M. We will use Method 2.9 to prove that M is defined by (5.14).

Let $\varphi : \{ b_i, c_i : i \in I \} \cup \{z\} \to M$ be given by $b_i \varphi = \tau_i$ and $c_i \varphi = \tau_i^{-1}$ for each $i \in I$, and $z\varphi = \varnothing$. Then for $i, j \in I$ with $i \neq j$,

$$(b_i c_i)\varphi^* = \tau_i \tau_i^{-1} = \mathrm{id}_X = \varepsilon\varphi^*; \tag{S.16}$$
$$(b_i c_j)\varphi^* = \tau_i \tau_j^{-1} = \varnothing = z\varphi^*; \tag{S.17}$$
$$(b_i z)\varphi^* = \tau_i \varnothing = \varnothing = z\varphi^*; \tag{S.18}$$
$$(z b_i)\varphi^* = \varnothing \tau_i = \varnothing = z\varphi^*; \tag{S.19}$$
$$(c_i z)\varphi^* = \tau_i^{-1} \varnothing = \varnothing = z\varphi^*; \tag{S.20}$$
$$(z c_i)\varphi^* = \varnothing \tau_i^{-1} = \varnothing = z\varphi^*; \tag{S.21}$$
$$(zz)\varphi^* = \varnothing\varnothing = \varnothing = z\varphi^*. \tag{S.22}$$

Thus M satisfies the defining relations in (5.14) with respect to φ. Let

$$N = \{ c_i : i \in I \}^* \{ b_i : i \in I \}^* \cup \{z\}.$$

Any word in $\{ z, b_i, c_i : i \in I \}^*$ can be transformed to one in N by applying defining relations to remove any subwords $b_i c_j$ (for any $i, j \in I$, replacing them with z if $i \neq j$), and then to replacing any two-symbol subword that contains a z into z alone.

The remaining step is to show that $\varphi^*|_N$ is injective. Now, $x \tau_i \tau_j^{-1}$ is defined if and only if $i = j$, and $x \tau_i^{-1}$ is defined if and only if $x \in \operatorname{im} \tau_i$. So

$$x \tau_{i_\ell} \tau_{i_{\ell-1}} \cdots \tau_{i_1} \tau_{j_1}^{-1} \tau_{j_2}^{-1} \cdots \tau_{j_m}^{-1}$$

is defined for all $x \in X$ if and only if $\ell \geqslant m$ and $i_h = j_h$ for $h = 1, \dots, m$.

So suppose

$$(c_{j_1} \cdots c_{j_m} b_{i_1} \cdots b_{i_n})\varphi^* = (c_{j'_1} \cdots c_{j'_{m'}} b_{i'_1} \cdots b_{i'_{n'}})\varphi^*.$$

Interchanging the two sides if necessary, assume $m \leqslant m'$. Now,

$$x\tau_{j_m}\tau_{j_{m-1}}\cdots\tau_{j_1}(c_{j_1}\cdots c_{j_m}b_{i_1}\cdots b_{i_n})\varphi^* = x\tau_{j_m}\tau_{j_{m-1}}\cdots\tau_{j_1}\tau_{j_1}^{-1}\cdots\tau_{j_m}^{-1}\tau_{i_1}\cdots\tau_{i_n}$$

is defined for all $x \in X$. So

$$x\tau_{j_m}\tau_{j_{m-1}}\cdots\tau_{j_1}(c_{j'_1}\cdots c_{j'_m}b_{i'_1}\cdots b_{i'_{n'}})\varphi^* = x\tau_{j_m}\tau_{j_{m-1}}\cdots\tau_{j_1}\tau_{j_1}^{-1}\cdots\tau_{j_m}^{-1}\tau_{i_1}\cdots\tau_{i_n}$$

is defined for all $x \in X$. So $m \geqslant m'$, and thus $m = m'$, and $j_h = j'_h$ for $h = 1, \ldots, m$. Now, $x = x\tau_{j_m}\tau_{j_{m-1}}\cdots\tau_{j_1}(c_{j_1}\cdots c_{j_m})\varphi^*$, so $x(b_{i_1}\cdots b_{i_n})\varphi^* = x(b_{i'_1}\cdots b_{i'_{n'}})\varphi^*$ for all $x \in X$. Interchanging the two sides if necessary, assume $n \geqslant n'$. Then

$$x(b_{i_1}\cdots b_{i_n})\varphi^*\tau_{i_n}^{-1}\cdots\tau_{i_1}^{-1} = x\tau_{i_1}\cdots\tau_{i_n}\tau_{i_n}^{-1}\cdots\tau_{i_1}^{-1}$$

is defined for all $x \in X$. So

$$x(b_{i'_1}\cdots b_{i'_{n'}})\varphi^*\tau_{i_n}^{-1}\cdots\tau_{i_1}^{-1} = x\tau_{i'_1}\cdots\tau_{i'_{n'}}\tau_{i_n}^{-1}\cdots\tau_{i_1}^{-1}$$

is defined for all $x \in X$. So $n \leqslant n'$, and thus $n = n'$, and $i_h = i'_h$ for $h = 1, \ldots, n$. Hence

$$c_{j_1}\cdots c_{j_m}b_{i_1}\cdots b_{i_n} = c_{j'_1}\cdots c_{j'_{m'}}b_{i'_1}\cdots b_{i'_{n'}}.$$

Finally, note that we have shown that $x(c_{j_1}\cdots c_{j_m}b_{i_1}\cdots b_{i_n})\varphi^*$ is always defined for some element $x \in X$. Hence $z\varphi^* = \varnothing \neq (c_{j_1}\cdots c_{j_m}b_{i_1}\cdots b_{i_n})\varphi^*$. Thus $\varphi^*|_N$ is injective.

5.6 a) Let S be a Clifford semigroup. Then $S \simeq \mathcal{S}[Y; G_\alpha; \varphi_{\alpha,\beta}]$, for some semilattice Y, groups G_α, and homomorphisms $\varphi_{\alpha,\beta} : G_\alpha \to G_\beta$. Let e and f be idempotents in S. Then $e \in G_\alpha$ and $f \in G_\beta$ for some $\alpha, \beta \in Y$. Thus $e = 1_\alpha$ and $f = 1_\beta$, where 1_α and 1_β are the identities of G_α and G_β. So $ef = 1_\alpha 1_\beta = (1_\alpha\varphi_{\alpha,\alpha\sqcap\beta})(1_\beta\varphi_{\beta,\alpha\sqcap\beta}) = 1_{\alpha\sqcap\beta}1_{\alpha\sqcap\beta} = 1_{\alpha\sqcap\beta}$. So the idempotents of S form a subsemigroup. Since S is regular by Theorem 5.13, S is orthodox.

b) Let S be completely simple and orthodox. By Theorem 4.11, $S \simeq \mathcal{M}[G; I, \Lambda; P]$ for some group G, index sets I and Λ and regular matrix P over G. View $I \times \Lambda$ as a rectangular band. Without loss of generality, assume that there is a symbol 1 in $I \cap \Lambda$. The elements $(1, p_{\lambda 1}^{-1}, \lambda)$ and $(j, p_{1j}^{-1}, 1)$ are idempotents of S, and so, since S is orthodox, their product $(1, p_{\lambda 1}^{-1}\lambda)(j, p_{1j}^{-1}, 1) = (1, p_{\lambda 1}^{-1}p_{\lambda j}p_{1j}^{-1}, 1)$ is also an idempotent; hence $p_{\lambda 1}^{-1}p_{\lambda j}p_{1j}^{-1} = p_{11}^{-1}$. Define a map $\varphi : G \times (I \times \Lambda) \to S$ by $(g, (i, \lambda))\varphi = (i, p_{1i}^{-1}gp_{11}p_{\lambda 1}^{-1}, \lambda)$. Then

$$\begin{aligned}&(g, (i, \lambda))\varphi(h, (j, \mu))\varphi\\ &= (i, p_{1i}^{-1}gp_{11}p_{\lambda 1}^{-1}, \lambda)(j, p_{1j}^{-1}hp_{11}p_{\mu 1}^{-1}, \mu)\end{aligned}$$

$$
\begin{aligned}
&= (i, p_{1i}^{-1} g p_{11} p_{\lambda 1}^{-1} p_{\lambda j} p_{1j}^{-1} h p_{11} p_{\mu 1}^{-1}, \mu) \\
&= (i, p_{1i}^{-1} g p_{11} p_{11}^{-1} h p_{11} p_{\mu 1}^{-1}, \mu) \\
&= (i, p_{1i}^{-1} g h p_{11} p_{\mu 1}^{-1}, \mu) \\
&= (gh, (i, \mu))\varphi;
\end{aligned}
$$

thus φ is a homomorphism. It is clearly injective and surjective and thus an isomorphism.

For the converse, let G be a group and let $I \times \Lambda$ be a rectangular band. Let P be the $\Lambda \times I$ matrix all of whose entries are 1_G. It is straightforward to see that

$$\varphi : G \times (I \times \Lambda) \to \mathcal{M}[G; I, \Lambda; P], \quad (g, (i, \lambda)) \mapsto (i, g, \lambda)$$

is an isomorphism; thus $G \times (I \times \Lambda)$ is completely simple. The only idempotent in G is 1_G and every element of $I \times \Lambda$ is idempotent. Hence the set of idempotents of $G \times (I \times \Lambda)$ is $\{1_G\} \times (I \times \Lambda)$, which is clearly a subsemigroup. Since $G \times (I \times \Lambda)$ is completely simple, it is regular by Proposition 4.13, and hence is orthodox.

5.7 Let S be a completely 0-simple inverse semigroup. By Theorem 4.7, $S \simeq \mathcal{M}_0[G; I, \Lambda; P]$ for some group G, index sets I and Λ, and regular matrix P over G^0. Since S is inverse, every $\mathcal{L}$-class and every $\mathcal{R}$-class contains exactly one idempotent. Now, the non-zero idempotents of $\mathcal{M}_0[G; I, \Lambda; P]$ are elements of the form $(i, p_{\lambda i}^{-1}, \lambda)$, where $i \in I$ and $\lambda \in \Lambda$ are such that $p_{\lambda i} \neq 0$. The non-zero $\mathcal{R}$-classes of $\mathcal{M}_0[G; I, \Lambda; P]$ are the sets $\{i\} \times G \times \Lambda$; the non-zero $\mathcal{L}$-classes of $\mathcal{M}_0[G; I, \Lambda; P]$ are the sets $I \times G \times \{\lambda\}$. So for each i, there is a unique λ such that $p_{\lambda i}$ is non-zero, and vice versa. Hence there is a bijection $\psi : I \to \Lambda$ so that $i\psi$ is the unique element of Λ with $p_{(i\psi)i} \neq 0$. Hence $|I| = |\Lambda|$ Since Λ an abstract index set, we can reorder it and the rows of P so that P becomes diagonal. Now we can simply replace the index set Λ with I.

Now suppose that $S \simeq \mathcal{M}_0[G; I, I; P]$, where P is diagonal. Then S is completely 0-simple and therefore regular. The idempotents of $\mathcal{M}_0[G; I, I; P]$ are the elements (i, p_{ii}^{-1}, i). If $i \neq j$, then $p_{ij} = 0$ (since P is diagonal) and so $(i, p_{ii}^{-1}, i)(j, p_{jj}^{-1}, j) = 0$. So the idempotents of S commute and so S is inverse.

5.8 a) Let $x \in \operatorname{im} \tau$. Then $x = z\tau$ for some $z \in S^1$. Let $y \in S^1$. Since τ is a partial right translation, $\operatorname{dom} \tau$ is a left ideal and so $yz \in \operatorname{dom} \tau$; furthermore, $(yz)\tau = y(z\tau) = yx$ and so $yx \in \operatorname{im} \tau$. Thus $\operatorname{im} \tau$ is a left ideal of S^1.

b) Let $\tau, \sigma \in \mathcal{I}_{S^1}$ be partial right translations. Let $x, y \in S^1$. Suppose $x\tau\sigma$ is defined. Then both $x \in \operatorname{dom} \tau$ and $x\tau \in \operatorname{dom} \sigma$. Since $\operatorname{dom} \tau$ is a left ideal, $yx \in \operatorname{dom} \tau$ and $(yx)\tau = y(x\tau)$. Since $\operatorname{dom} \sigma$ is a left ideal, $y(x\tau) \in \operatorname{dom} \sigma$ and $(y(x\tau))\sigma = y(x\tau\sigma)$. Hence $yx \in \operatorname{dom}(\tau\sigma)$ and $(yx)\tau\sigma = y(x\tau\sigma)$. So $\tau\sigma$ is a partial right translation.

Suppose $x\tau^{-1}$ is defined. Let $z = x\tau^{-1}$. Then $z \in \operatorname{dom}\tau$ and $z\tau = x$. Since $\operatorname{dom}\tau$ is a left ideal, $yz \in \operatorname{dom}\tau$ and $(yz)\tau = y(z\tau) = yx$. So $yx \in \operatorname{dom}\tau^{-1}$ and $(yx)\tau^{-1} = yz = y(x\tau^{-1})$. So τ^{-1} is a partial right translation.

Hence the set of partial right translations forms an inverse subsemigroup of $\mathcal{I}_{S^1}$. Since every ρ_x is a partial right translation, T is a subsemigroup of the set of partial right translation.

5.9 Let $x = c^\gamma b^\beta \in B$ be arbitrary. Let $y = c^\beta b^\gamma$. Then

$$xyx = c^\gamma b^\beta c^\beta b^\gamma c^\gamma b^\beta =_B c^\gamma b^\beta = c^\gamma b^\beta = x.$$

So x is regular. The idempotents of B are elements of the form $c^\gamma b^\gamma$ by Exercise 2.10(a). Thus, given two idempotents $e = c^\gamma b^\gamma$ and $f = c^\beta b^\beta$, we see that if $\gamma \geqslant \beta$,

$$\begin{aligned} ef = c^\gamma b^\gamma c^\beta b^\beta =_B c^\gamma b^{\gamma-\beta} b^\beta =_B c^\gamma c^\gamma \\ =_B c^\beta c^{\gamma-\beta} b^\gamma =_B c^\beta b^\beta c^\gamma b^\gamma = fe \end{aligned}$$

and similarly $ef = fe$ if $\gamma \leqslant \beta$. So B is a regular semigroup whose idempotents commute and so is inverse by Theorem 5.1.

5.10 For $x \in S$ and $e \in E(S)$,

$$\begin{aligned} x \preccurlyeq e &\Rightarrow x = xx^{-1}e && \text{[by definition of } \preccurlyeq\text{]} \\ &\Rightarrow x^2 = xx^{-1}exx^{-1}e \\ &\Rightarrow x^2 = xx^{-1}xx^{-1}ee && \text{[since idempotents commute in } S\text{]} \\ &\Rightarrow x^2 = xx^{-1}e && \text{[since } xx^{-1} \text{ and } e \text{ are idempotents]} \\ &\Rightarrow x^2 = x && \text{[since } x = xx^{-1}e\text{]} \\ &\Rightarrow x \in E(S). \end{aligned}$$

5.11 Consider an element u of $\mathrm{FInvM}(\{\alpha\})$. Let T_1 be the (unique) Munn tree corresponding to u. Let p, q, and r be, respectively, the 'x-coordinates' of the leftmost endpoint, the vertex ω_T, and the rightmost endpoint. Notice that $p \leqslant 0, r \geqslant 0$, and $p \leqslant q \leqslant r$, so that $(p, q, r) \in K$. In this way, we determine a map $\varphi : \mathrm{FInvM}(\{a\}) \to K$. Clearly, a unique Munn tree of the given form can be reconstructed from any triple $(p, q, r) \in K$, so φ is injective and surjective.

Let v be another element of $\mathrm{FInvM}(\{a\})$ and let T_2 be the corresponding Munn tree. Let the triple $(p', q', r') \in K$ correspond to T_2. Consider multiplying u and v using the corresponding Munn trees T_1 and T_2 to get a Munn tree T corresponding uv. The process is illustrated in Figure S.5. First we merge the vertices ω_{T_1} and α_{T_2} to form a vertex that we call ∞, and let $\alpha_T = \alpha_{T_1}$ and $\omega_T = \omega_{T_2}$. Then we fold edges together until we get the Munn tree T. It is easy to see from the diagram that the coordinate of ω_T relative to α_T is $q + q'$, that the coordinate of the leftmost endpoint of T relative to α_T is the

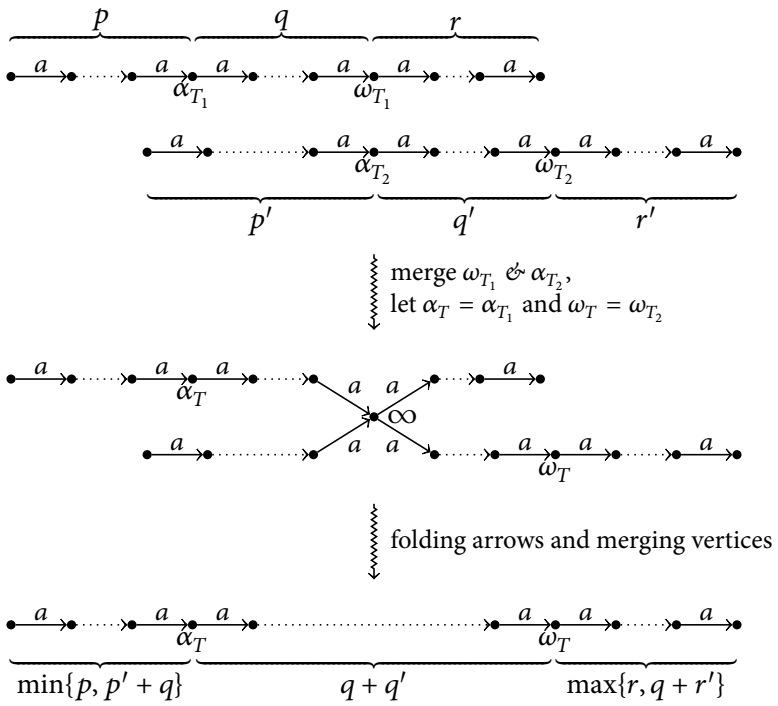


FIGURE S.5
Multiplication of elements of $\mathrm{FInvM}(\{a\})$ using Munn trees. The numbers p, q, r are, respectively, the 'x-coordinates' relative to α_{T_1} of the left endpoint of T_1, the vertex ω_{T_1}, and the right endpoint of T_1; the numbers p', q', r' play a similar role for T_2.

smaller of p and $q + p'$, and the coordinate of the rightmost endpoint of T relative to α_T is the greater of r and $q + r'$. That is, the triple $(\min\{p, q+p'\}, q+q', \max\{r, q+r'\})$ corresponds to T. Thus the map φ is a homomorphism and thus an isomorphism.

5.12 By Exercise 5.11, the monoid K is isomorphic to $\mathrm{FInvM}(\{a\})$. Thus it is sufficient to prove that K is a subdirect product of $B \times B$. The map φ is a homomorphism since

$$\begin{aligned}
&(p,q,r)\varphi(p',q',r')\varphi \\
&= (c^{-p}b^{-p+q}, c^{r}b^{-q+r})(c^{-p'}b^{-p'+q'}, c^{r'}b^{-q'+r'}) \\
&= (c^{-p+p-q+\max\{-p+q,-p'\}}b^{-p'+q'+p'+\max\{-p+q,-p'\}}, \\
&\qquad\qquad c^{r+q-r+\max\{-q+r,r'\}}b^{r'-q'-r'+\max\{-q+r,r'\}}) \\
&= (c^{\max\{-p,-p'-q\}}b^{\max\{-p+q+q',-p'+q'\}}, c^{\max\{r,q+r'\}}b^{\max\{r-q-q',-q'+r'\}}) \\
&= (c^{\max\{-p,-p'-q\}}b^{q+q'+\max\{-p,-p'-q\}}, c^{\max\{r,q+r'\}}b^{\max\{r,q+r'\}-q-q'}) \\
&= (-\max\{-p,-p'-q\}, q+q', \max\{r,q+r'\})\varphi \\
&= (\min\{p,p'+q\}, q+q', \max\{r,q+r'\})\varphi \\
&= \big((p,q,r)(p',q',r')\big)\varphi.
\end{aligned}$$

Furthermore, φ is injective since

$$\begin{aligned}
&(p,q,r)\varphi = (p',q',r')\varphi \\
&\Rightarrow (c^{-p}b^{-p+q}, c^{r}b^{-q+r}) = (c^{-p'}b^{-p'+q'}, c^{r'}b^{-q'+r'})
\end{aligned}$$

$$\Rightarrow (-p = -p') \wedge (-p + q = -p' + q') \wedge (r = r')$$
$$\Rightarrow (p, q, r) = (p', q', r').$$

So φ embeds K into $B \times B$. Finally, as p and q range over $\mathbb{N} \cup \{0\}$, clearly $(p, q, q)\varphi\pi_1 = c^{-p}b^{-p+q}$ ranges over B, and as q and r range over $\mathbb{N} \cup \{0\}$, clearly $(q, q, r)\varphi\pi_2 = c^r b^{-q+r}$ ranges over B. So $\operatorname{im}\varphi$ projects surjectively to both copies of B, and so K is a subdirect product of two copies of B.

5.13 a) Since $\mathrm{BR}(M, \varphi)$ is generated by $A \cup \{b, c\}$, every element is represented by some word $u \in (A \cup \{b, c\})^*$. Using the defining relations (bc, ε), we can delete any subword bc. Then, using defining relations of the form $(ba, (a\varphi)b)$, we can replace any subword ba by $(a\varphi)b$ and any subword ac by $c(a\varphi)$. Iterating this process, we eventually find a word v containing no subwords bc, ba or ac: that is, $v = c^\gamma w b^\beta$ for some $\gamma, \beta \in \mathbb{N} \cup \{0\}$ and $w \in A^*$.

b) i) Suppose that $\gamma = \gamma'$, $\beta = \beta'$, and $w =_M w'$. Then there is a sequence of elementary ρ-transitions from w to w'. Since ρ is a subset of the defining relations in (5.15), w and w' represent the same element of $\mathrm{BR}(M, \varphi)$. Hence $c^\gamma w b^\beta$ and $c^\gamma w' b^\beta$ represent the same element of $\mathrm{BR}(M, \varphi)$.

ii) It is easy to prove that for all defining relations (u, v) in (5.15), we have $u\psi = v\psi$ and so ψ is well-defined.

Suppose now that $c^\gamma w b^\beta$ and $c^{\gamma'} w' b^{\beta'}$ represent the same element of $\mathrm{BR}(M, \varphi)$. Then $(c^\gamma w b^\beta)\psi = (c^{\gamma'} w' b^{\beta'})\psi$. Thus

$$\begin{aligned}(\gamma, w, \beta) &= (0, 1_M, 0)\big((c^\gamma w b^\beta)\psi\big)\\ &= (0, 1_M, 0)\big((c^{\gamma'} w' b^{\beta'})\psi\big) = (\gamma, w', \beta),\end{aligned}$$

and so $\gamma = \gamma'$ and $\beta = \beta'$.

c) Define a map $\vartheta : M \to \mathrm{BR}(M, \varphi)$ by $w\vartheta = w$. This is clearly a homomorphism, and

$$w\varphi =_{\mathrm{BR}(M,\varphi)} w'\varphi \Rightarrow c^0 w b^0 =_{\mathrm{BR}(M,\varphi)} c^0 w' b^0 \Rightarrow w =_M w'$$

by parts a) and b). Hence ϑ is injective and so M embeds into $\mathrm{BR}(M, \varphi)$.

5.14 Let $S = \mathrm{BR}(M, \varphi)$. We aim to show that $SxS = S$ for all $x \in S$. Suppose $x = c^\gamma w b^\beta$, where $w \in M$. Let $c^\delta u b^\zeta$ be an arbitrary element of S. Let $p = c^\delta u b^{\gamma+1}$ and $q = c^{\beta+1} b^\zeta$. Then

$$\begin{aligned}pxq &= c^\delta u b^{\gamma+1} c^\gamma w b^\beta c^{\beta+1} b^\zeta\\ &=_S c^\delta u b w c b^\zeta\\ &=_S c^\delta u b c (w\varphi) b^\zeta\\ &=_S c^\delta u b c b^\zeta\\ &=_S c^\delta u b^\zeta.\end{aligned}$$

So $c^\delta u b^\zeta = pxq \in SxS$. Since $c^\delta u b^\zeta \in S$ was arbitrary, $S = SxS$. Hence any ideal of S must be S itself. So S is simple.

EXERCISES FOR CHAPTER 6

[See page 128 for the exercises.]

6.1 For clarity, let $\iota : S \to G$ and $\iota' : S \to G'$ be the embedding maps. Define $\psi : G \to H$ by $(x\iota)(y\iota)^{-1}\varphi = (x\iota')(y\iota')^{-1}$ for $x, y \in S$. Let $x_1, x_2, y_1, y_2 \in S$. Then

$$\begin{aligned}
&(x_1\iota)(y_1\iota)^{-1} = (x_2\iota)(y_2\iota)^{-1} \\
\Leftrightarrow\ &(x_1\iota)(y_2\iota) = (x_2\iota)(y_1\iota) \\
\Leftrightarrow\ &x_1y_2 = x_2y_1 && \text{[since } \iota \text{ is an injective homomorphism]} \\
\Leftrightarrow\ &(x_1\iota')(y_2\iota') = (x_2\iota')(y_1\iota') \\
& && \text{[since } \iota' \text{ is an injective homomorphism]} \\
\Leftrightarrow\ &(x_1\iota')(y_1\iota')^{-1} = (x_2\iota')(y_2\iota')^{-1} \\
\Leftrightarrow\ &\big((x_1\iota)(y_1\iota)^{-1}\big)\psi = \big((x_2\iota)(y_2\iota)^{-1}\big)\psi.
\end{aligned}$$

The forward implication shows ψ is well-defined; the reverse implication shows it is injective. Furthermore

$$\begin{aligned}
&\big((x_1\iota)(y_1\iota)^{-1}\big)\psi\big((x_2\iota)(y_2\iota)^{-1}\big)\psi \\
&= (x_1\iota')(y_1\iota')^{-1}(x_2\iota')(y_2\iota')^{-1} && \text{[by definition of } \psi\text{]} \\
&= (x_1\iota')(x_2\iota')(y_1\iota')^{-1}(y_2\iota')^{-1} && \text{[by commutativity]} \\
&= (x_1x_2)\iota'((y_2y_1)\iota')^{-1} && \text{[by inverses in } H\text{]} \\
&= \big(((x_1x_2)\iota)(y_2y_1\iota)^{-1}\big)\psi && \text{[by definition of } \psi\text{]} \\
&= \big((x_1\iota)(x_2\iota)(y_1\iota)^{-1}(y_2\iota)^{-1}\big)\psi && \text{[by inverses in } G\text{]} \\
&= \big(\big((x_1\iota)(y_1\iota)^{-1}\big)\big((x_2\iota)(y_2\iota)^{-1}\big)\big)\psi, && \text{[by commutativity]}
\end{aligned}$$

so ψ is a homomorphism.

Finally, let $s\iota \in S\iota$. Then for arbitrary $z \in S$,

$$(s\iota)\psi = ((sz\iota)(z\iota)^{-1})\psi = ((sz\iota')(z\iota')^{-1}) = s\iota',$$

so ψ is clearly maps $S\iota$ surjectively to $S\iota'$.

6.2 Fix $x \in I$. For $s \in S \smallsetminus I$. Define $s\widehat{\varphi}$ to be $(x\varphi)^{-1}((xs)\varphi)$; notice that $xs \in I$ since I is an ideal. Now, for $s' \in S$ and $y \in I$,

$$\begin{aligned}
&(s\widehat{\varphi})(y\widehat{\varphi}) \\
&= (x\varphi)^{-1}((xs)\varphi)(y\varphi) && \text{[by definition of } \widehat{\varphi}\text{]} \\
&= (x\varphi)^{-1}((xsy)\varphi) && \text{[since } \varphi \text{ is a homomorphism]} \\
&= (sy)\widehat{\varphi}; && \text{[by definition of } \widehat{\varphi}\text{]}
\end{aligned}$$

furthermore, $(s\widehat{\varphi})(y\widehat{\varphi}) = (sy)\widehat{\varphi}$ by commutativity of S and G. For $s, s' \in S$,

$$\begin{aligned}
&(s\widehat{\varphi})(s'\widehat{\varphi}) \\
&= (x\varphi)^{-1}((xs)\varphi)(x\varphi)^{-1}((xs')\varphi) && \text{[by definition of } \widehat{\varphi}] \\
&= (x\varphi)^{-1}(x\varphi)^{-1}((xs)\varphi)((xs')\varphi) && \text{[since } G \text{ is abelian]} \\
&= (x\varphi)^{-1}(x\varphi)^{-1}((xsxs')\varphi) && \text{[since } \varphi \text{ is a homomorphism]} \\
&= (x\varphi)^{-1}(x\varphi)^{-1}((xxss')\varphi) && \text{[since } S \text{ is commutative]} \\
&= (x\varphi)^{-1}(x\varphi)^{-1}(x\varphi)((xss')\varphi) \\
& && \text{[since } \varphi \text{ is a homomorphism and } x, xss' \in I] \\
&= (x\varphi)^{-1}((xss')\varphi) && \text{[since } (x\varphi)^{-1}(x\varphi) = 1_G] \\
&= (ss')\widehat{\varphi}. && \text{[by definition of } \widehat{\varphi}]
\end{aligned}$$

Together with the fact that φ is a homomorphism, this shows that $\widehat{\varphi}$ is a homomorphism.

Finally, suppose $\psi : S \to G$ is a homomorphism extending φ. Then $(xs)\psi = (x\psi)(s\psi)$ for any $s \in S \setminus I$. Hence $(xs)\varphi = (x\varphi)(s\psi)$ since $x, xs \in I$, and so $s\psi = (x\varphi)^{-1}((xs)\varphi) = s\widehat{\varphi}$. Hence $\psi = \widehat{\varphi}$ and so $\widehat{\varphi}$ is the unique extension of φ to S.

6.3 Let $d = \gcd(S)$; this is well-defined since $S \neq \{0\}$. Then if $x \in S$, then $x = dk \in d\mathbb{N}$, so $S \subseteq d\mathbb{N}$. Furthermore, there exist $z_1, \ldots, z_n \in S$ and $k_1, \ldots, k_n \in \mathbb{Z}$ such that $k_1z_1 + k_2z_2 + \cdots + k_nz_n = d$, hence moving all the terms where k_i is negative to the right of the equality, we get $s + d = s'$ for two elements s and s' of S. Suppose $s = dt$ and $s' = dt'$. Now let $n \in \mathbb{N}$ with $n \geqslant (t-1)t + (t-1)$; we aim to prove that $dn \in S$. Let $n = qt + r$, where $q \in \mathbb{N}$ and $0 \leqslant r < t$. Then $q \geqslant (t-1) \geqslant r$ and so $q - r \geqslant 0$. Now, $n = (rt + r) + (q - r)t = r(t + 1) + (q - r)t$, and so $dn = r(dt + d) + (q - r)td = r(s + d) + (q - r)s = rs' + (q - r)s \in S$. Thus, if $x \in d\mathbb{N} \setminus S$, then $x = dn$ for $n < (t-1)t + (t-1)$. Thus $d\mathbb{N} \setminus S$ is finite.

6.4 If $S = \{0\}$; then all three conditions hold. So assume $S \neq \{0\}$ and suppose S contains both a positive integer p and a negative integer n. Let $S_+ = \{s \in S : s > 0\}$ and $S_- = \{s \in S : s < 0\}$; clearly S_+ and S_- are subsemigroups of S. Let $d = \gcd(S)$, $d_+ = \gcd(S_+)$, and $d_- = \gcd(S_-)$. Clearly, $d \leqslant d_+$ and $d \leqslant d_-$. Since $d = s - s'$ for some $s, s' \in S$, we have $d = (s + kp) - (s' + kp) = (s + kn) - (s' + kn)$ for all $k \in \mathbb{N}$. Thus d is both the difference between two elements of S_+ and the difference between two elements of S_-. Hence $d \geqslant d_+$ and $d \geqslant d_-$ and hence $d = d_+ = d_-$. Thus $S_+ \subseteq d\mathbb{N}$ and $S_- \subseteq -d\mathbb{N}$, and $d\mathbb{N} \setminus S_+$ and $-d\mathbb{N} \setminus S_-$ are finite. Hence $dk, d(k+2) \in S_+ \subseteq S$ and $-d(k+1) \in S_- \subseteq S$ for large k. Hence $d = d(k+2) - d(k+1) \in S$ and $-d = dk - d(k+1) \in S$. So $d\mathbb{Z} \subseteq S \subseteq S_- \cup \{0\} \cup S_+ \subseteq -d\mathbb{N} \cup \{0\} \cup d\mathbb{N} = d\mathbb{Z}$. Hence $S = d\mathbb{Z}$ is a subgroup of $\mathbb{Z}$.

6.5 a) From the definition, $\sim$ is clearly reflexive and symmetric. Suppose $\alpha \sim \beta$ and $\beta \sim \gamma$. Then there exist δ and ζ with $\delta \subseteq \alpha$, $\delta \subseteq \beta$, $\zeta \subseteq \alpha$, and $\zeta \subseteq \beta$. Let $\eta = \zeta\zeta^{-1}\delta$. Then $\operatorname{dom}\eta \subseteq \operatorname{dom}\zeta$ and for any $x \in \operatorname{dom}\eta$, we have

$$x\eta = x\zeta\zeta^{-1}\delta = x\delta = x\beta = x\zeta$$

and so $\eta \subseteq \delta \subseteq \alpha$ and $\eta \subseteq \zeta \subseteq \gamma$. Hence $\alpha \sim \gamma$. Therefore $\sim$ is transitive.

Suppose $\alpha_1 \sim \beta_1$ and $\alpha_2 \sim \beta_2$. Then there exist δ_1 and δ_2 with $\delta_1 \subseteq \alpha_1$, $\delta_1 \subseteq \beta_1$, $\delta_2 \subseteq \alpha_2$ and $\delta_2 \subseteq \beta_2$. Hence $\delta_1\delta_2 \subseteq \alpha_1\alpha_2$ and $\delta_1\delta_2 \subseteq \beta_1\beta_2$. Hence $\alpha_1\alpha_2 \sim \beta_1\beta_2$. Therefore $\sim$ is a congruence.

b) Let $\alpha, \beta \in T$. Let $\zeta = \alpha^{-1}\beta$ and $\eta = \beta\alpha^{-1}$. Then $\alpha\zeta = \alpha\alpha^{-1}\beta \subseteq \beta$ and so $\alpha\zeta \sim \beta$; similarly $\eta\alpha = \beta\alpha^{-1}\alpha \subseteq \beta$ and so $\eta\alpha \sim \beta$. Thus for any $[\alpha]_\sim, [\beta]_\sim \in G$, there exist $[\zeta]_\sim, [\eta]_\sim \in G$ with $[\alpha]_\sim[\zeta]_\sim = [\eta]_\sim[\alpha]_\sim = [\beta]_\sim$; hence $[\alpha]_\sim G = G[\alpha]_\sim = G$ for any $[\alpha]_\sim \in G$. Thus G is a group.

c) Let $\alpha, \beta \in T$. Then $\operatorname{im}\alpha$ is a left ideal of S by Exercise 5.8(a) and $\operatorname{dom}\beta$ is a left ideal of S since β is a partial right transformation. Since S is right-reversible, $\operatorname{im}\alpha \cap \operatorname{dom}\beta \neq \varnothing$. Hence $\alpha\beta \neq \varnothing$.

Since T is generated by the non-empty elements ρ_x and ρ_x^{-1}, we see that T does not contain the empty relation.

d) Suppose $x\psi = y\psi$; then $[\rho_x]_\sim = [\rho_y]_\sim$ and so $\rho_x \sim \rho_y$. Then there exists $\delta \in T$ such that $\delta \subseteq \rho_x$ and $\delta \subseteq \rho_y$. By the previous paragraph, δ is not the empty relation. So let $z \in \operatorname{dom}\delta$. Then $z\rho_x = z\rho_y$. Thus $zx = zy$ and so $x = y$ by cancellativity. Hence $\psi : S \to G$ is a monomorphism and so S is group-embeddable.

6.6 Let $(m,n), (p,q), (r,s) \in S$. Then

$$\begin{aligned}(m,n)\big((p,q)(r,s)\big) &= (m,n)(p+r, 2^r q + s)\\ &= (m+p+r, 2^{p+r}n + 2^r q + s)\\ &= (m+p+r, 2^r(2^p n + q) + s)\\ &= (m+p, 2^p n + q)(r,s)\\ &= \big((m,n)(p,q)\big)(r,s);\end{aligned}$$

thus the multiplication is associative.

Let $(m_1,n_1), (m_2,n_2) \in S$. Let $p_1 = m_2$, $q_1 = 2^{m_1}n_2$, $p_2 = m_1$, and $q_2 = 2^{m_2}n_2$. Then

$$\begin{aligned}(m_1,n_1)(p_1,q_1) &= (m_1 + p_1, 2^{p_1}n_1 + q_1)\\ &= (m_1 + m_2, 2^{m_2}n_2 + 2^{m_1}n_2)\end{aligned}$$

and

$$\begin{aligned}(m_2,n_2)(p_2,q_2) &= (m_2 + p_2, 2^{p_2}n_2 + q_2)\\ &= (m_2 + m_1, 2^{m_1}n_2 + 2^{m_2}n_2);\end{aligned}$$

so $(m_1, n_1)(p_1, q_1) = (m_2, n_2)(p_2, q_2)$. Since (m_1, n_1) and (m_2, n_2) were arbitrary, S is left-reversible.

Suppose S is right-reversible. Then $(1, 0)$ and $(1, 1)$ have a common left multiple. Hence there exist elements (p_1, q_1) and (p_2, q_2) such that $(p_1, q_1)(1, 0) = (p_2, q_2)(1, 1)$. Thus $(p_1 + 1, 2q_1) = (p_2 + 1, 2q_2 + 1)$, which is a contradiction, since $2q_1$ is even and $2q_2 + 1$ is odd. Therefore S is not right-reversible.

EXERCISES FOR CHAPTER 7

[See pages 147–148 for the exercises.]

7.1 Let M be a group. Then M is simple and so $MxM = M$ for all $x \in M$.

Now suppose $MxM = M$ for all $x \in M$. Then for each $x \in M$, there exists $p, q \in M$ such that $pxq = 1_M$. Hence $x \mathrel{\mathcal{J}} 1_M$ and so $x \mathrel{\mathcal{H}} 1_M$ by Proposition 7.1. Thus x lies in the group of units of M. So all elements of M are invertible and so M is a group.

7.2 In finite semigroups, $\mathcal{J} = \mathcal{D}$, so $J_x = D_x$. Since D_x is non-trivial, it contains some element $z \neq x$ such that $z \mathrel{\mathcal{R}} x$. That is, there exist $p, q \in S^1$ such that $xp = z$ and $zq = x$; notice that $p, q \in S$ since $x \neq z$. Hence $xpq = x$, and so $x(pq)^k = x$ for all $k \in \mathbb{N}$. Since S is finite, there is some $\ell \in \mathbb{N}$ such that $(pq)^\ell$ is idempotent. Let $y = (pq)^\ell$; then $y^2 = y$ and $xy = x$. By the ordering of $\mathcal{J}$-classes, $J_x = J_{xy} \leqslant J_y$. Since y is idempotent and thus regular, every element of $D_y = J_y$ is regular by Proposition 3.19.

7.3 a) Let S be a finite nilsemigroup. Let $n = |S|$. Let $x_1, \ldots, x_{n+1} \in S$. Consider the $n + 1$ products

$$x_1, \quad x_1x_2, \quad \ldots, \quad x_1 \cdots x_n, \quad x_1 \cdots x_{n+1}.$$

Since $|S| = n$, at least two of these $n + 1$ products must be equal: that is, $x_1 \cdots x_k = x_1 \cdots x_{k+\ell}$ for some $k \in \{1, \ldots, n\}$ and $\ell \in \{1, \ldots, n + 1 - k\}$. Hence

$$x_1 \cdots x_k = x_1 \cdots x_k x_{k+1} \cdots x_{k+\ell} = x_1 \cdots x_k (x_{k+1} \cdots x_{k+\ell})^m$$

for all $m \in \mathbb{N}$. Since S is a nilsemigroup, there is some $m \in \mathbb{N}$ with $(x_{k+1} \cdots x_{k+\ell})^m = 0$. Thus $x_1 \cdots x_k = x_1 \cdots x_k (x_{k+1} \cdots x_{k+\ell})^m = 0$ and so $x_1 \cdots x_n = 0$ (since $k \leqslant n$). Therefore $S^n = \{0\}$ and so S is nilpotent.

b) Let $S = \{0\} \cup \{ x_{i,j} : i \in \mathbb{N}, j \leqslant i \}$. Define a product on S as follows:

$$x_{i,j}x_{k,\ell} = \begin{cases} x_{i,j+\ell} & \text{if } i = k \text{ and } j + \ell \leqslant i, \\ 0 & \text{otherwise,} \end{cases}$$

$$x_{i,j}0 = 0x_{i,j} = 00 = 0.$$

It is easy to check that this operation is associative. For any $x_{i,j} \in S$, we have $x_{i,j}^{i+1} = 0$ since $j(i+1) > i$. Thus S is a nilsemigroup. However, for any $n \in \mathbb{N}$, we have $x_{n,1}^n = x_{n,n} \neq 0$, so $S^n \neq \{0\}$. Thus S is not nilpotent.

7.4 a) Let $x', y' \in J\varphi$. Then $x' = x\varphi$ and $y' = y\varphi$ for some $x, y \in J$. Thus there exist $p, q, r, s \in S^1$ such that $pxq = y$ and $rys = x$. Then $(p\varphi)x'(q\varphi) = y'$ and $(r\varphi)y'(s\varphi) = x'$ (where we view 1φ as the identity of $(S')^1$) and so $x' \mathcal{J} y'$. So all elements of $J\varphi$ are contained within a single $\mathcal{J}$-class J' of S'.

b) Let $x' \in J'$. Then $x' = x\varphi$ for some $x \in S$. Let $J = J_x$. Since all elements of $J\varphi$ are $\mathcal{J}$-related by part a), we see that $J\varphi \subseteq J'$.

Let J be minimal such that $J\varphi \subseteq J'$. Let $I = S^1JS^1$. Then $I = S^1xS^1$ for any $x \in J$, by the definition of $\mathcal{J}$. Let $y' \in J'$. Then $y' \mathcal{J} x\varphi$ and so there exist $p', q' \in (S')^1$ such that $y' = p'(x\varphi)q'$. Therefore $y' \in (S')^1(x\varphi)(S')^1 = (S^1xS^1)\varphi = I\varphi$ since φ is surjective. So $J' \subseteq I\varphi$.

Let $y \in I$ and let $K = J_y$. By part a), there exists some $\mathcal{J}$-class K' of S' such that $K\varphi \subseteq K'$. We now want to prove that $y \notin J$ implies $y\varphi \notin J'$. So suppose that $y \notin J$. Then $K = J_y < J$. Therefore $K\varphi \nsubseteq J'$ since J was chosen to be minimal such that $J\varphi \subseteq J'$. Hence $K' \neq J'$. Suppose, with the aim of obtaining a contradiction, that $y\varphi \in J'$. Then there exists $p', q', r', s' \in (S')^1$ with $p'(y\varphi)q' = x\varphi$ and $r'(x\varphi)s' = y\varphi$ for some $x \in J$. Since φ is surjective, this shows that $y \mathcal{J} x$ and so $y \in J$, which is a contradiction. Therefore $y\varphi \notin J'$.

Thus for any $y \in I$, we have $y \notin J$ implies $y\varphi \notin J'$. Hence $y\varphi \in J'$ implies $y \in J$, which implies $y\varphi \in J\varphi$. Since $J' \subseteq I\varphi$, this shows that $J' \subseteq J\varphi$. Thus $J\varphi = J'$.

7.5 It suffices to prove this when T is a subsemigroup of S and when T is a homomorphic image of S. In both cases, T is finite because S is, and thus for both S and T the property of having $\mathcal{H}$ being the equality relation is equivalent to aperiodic.

Let T be a subsemigroup of S. Let $x \in T$. Since $x \in S$ and S is aperiodic, there exists $k \in \mathbb{N}$ such that $x^k = x^{k+1}$. Since this is true for all $x \in T$, the subsemigroup T is aperiodic. Now let $\varphi : S \to T$ be a surjective homomorphism. Let $y \in T$. Then there exists $x \in S$ such that $x\varphi = y$. Since S is aperiodic, $x^k = x^{k+1}$ for some $k \in \mathbb{N}$. Hence $y^k = (x\varphi)^k = x^k\varphi = x^{k+1}\varphi = (x\varphi)^{k+1} = y^{k+1}$. Since this is true for all $y \in T$, the semigroup T is aperiodic. This completes the proof.

In the free semigroup $\{a\}^+$, the relation $\mathcal{H}$ is the equality relation, but any finite non-trivial cyclic group is a homomorphic image of $\{a\}^+$, and in groups all elements are $\mathcal{H}$-related.

7.6 Let $(s_1,t_1),(s_2,t_2),(s_3,t_3) \in S \rtimes_\varphi T$. Then

$$\begin{aligned}
&((s_1,t_1)(s_2,t_2))(s_3,t_3)\\
&= (s_1\,{}^{t_1}s_2, t_1t_2)(s_3,t_3) && \text{[by (7.1)]}\\
&= (s_1\,{}^{t_1}s_2\,{}^{t_1t_2}s_3, t_1t_2t_3) && \text{[by (7.1)]}\\
&= (s_1\,{}^{t_1}s_2\,{}^{t_1({}^{t_2}s_3)}, t_1t_2t_3) && \text{[by the definition of a left action]}\\
&= (s_1\,{}^{t_1}(s_2\,{}^{t_2}s_3), t_1t_2t_3) && \text{[since the action is by endomorphisms]}\\
&= (s_1,t_1)(s_2\,{}^{t_2}s_3, t_2t_3) && \text{[by (7.1)]}\\
&= (s_1,t_1)((s_2,t_2)(s_3,t_3)); && \text{[by (7.1)]}
\end{aligned}$$

thus the multiplication (7.1) is associative.

7.7 Suppose M and N are groups. Then $M \wr N$ is a monoid with identity $(e,1_N)$ by Proposition 7.7. Let $(f,n) \in M \wr N$. Define $f' \in N \to M$ by $(x)f' = ((xn^{-1})f)^{-1}$. Then

$$\begin{aligned}
&(f,n)(f',n^{-1})\\
&= (f\,{}^nf', nn^{-1})\\
&= (e,1_N),
\end{aligned}$$

since

$$\begin{aligned}
(x)f\,{}^nf' &= (x)f(xn)f' = (x)f((xnn^{-1})f)^{-1}\\
&= (x)f((x)f)^{-1} = 1_M,
\end{aligned}$$

and

$$\begin{aligned}
&(f',n^{-1})(f,n)\\
&= (f'\,{}^{n^{-1}}f, n^{-1}n)\\
&= (e,1_N),
\end{aligned}$$

since

$$(x)f'\,{}^{n^{-1}}f = (x)f'(xn^{-1})f = (x)f'((x)f')^{-1} = 1_M;$$

thus (f',n^{-1}) is a right and left inverse for (f,n). Hence $M \wr N$ is a group.

7.8 The wreath product $S \wr T$ must be right-cancellative but is not necessarily left-cancellative. For $(f,s),(g,t),(h,u) \in S \wr T$,

$$\begin{aligned}
&(f,s)(h,u) = (g,t)(h,u)\\
&\Rightarrow (f\,{}^sh, su) = (g\,{}^th, tu)\\
&\Rightarrow f\,{}^sh = g\,{}^th \wedge su = tu\\
&\Rightarrow (\forall x \in T)((x)f(xs)h = (x)g(xt)h) \wedge s = t
\end{aligned}$$

[since T is cancellative]

$$\Rightarrow (\forall x \in T)((x)f(xs)h = (x)g(xs)h) \wedge s = t$$

[substituting $s = t$]

$$\Rightarrow (\forall x \in T)((x)f = (x)g) \wedge s = t \qquad \text{[since } S \text{ is cancellative]}$$

$$\Rightarrow f = g \wedge s = t.$$

Now let $S = T = \mathbb{N} \cup \{0\}$ (under $+$) and define a map $f : S \to T$ by $(0)f = 1$ and $(x)f = 0$ for all $x \in T \setminus \{0\}$ and a map $g : S \to T$ by $(x)g = 0$ for all $x \in T$. Then

$$(g, 1)(f, 1) = (g\,{}^1f, 2) = (g\,{}^1g, 2) = (g, 1)(g, 1)$$

since $(x)g\,{}^1f = (x)g + (x+1)f = 0 + 0 = (x)g + (x+1)g = (x)g\,{}^1g$ for all $x \in T$. Hence $S \wr T$ is not left-cancellative.

7.9 This is a tedious analysis of products of three elements in $C(S)$. Each element is either in S or S'; there are thus eight cases. Let $x, y, z \in S$. Then:

- $(xy)z = x(yz)$, since S is a subsemigroup of $S \cup S'$;
- $(xy)z' = z' = xz' = x(yz')$;
- $(xy')z = y'z = (yz)' = x(yz)' = x(y'z)$;
- $(xy')z' = z' = xz' = x(y'z')$;
- $(x'y)z = (xy)'z = ((xy)z)' = (x(yz))' = x'(yz)$, using associativity in S for the third equality;
- $(x'y)z' = z' = x'z' = x'(yz')$;
- $(x'y')z = y'z = (yz)' = x'(yz)' = x'(y'z)$;
- $(x'y')z' = z' = x'z^{'} = x'(y'z')$.

Therefore the product defined by (7.4) is associative.

7.10 Define a map $\psi : C(M) \to \mathcal{T}_M$ by $x\psi = \rho_x$ and $x'\psi = \tau_x$ for $x \in M$. Clearly $\operatorname{im}\varphi = \{\rho_x, \tau_x : x \in M\}$. We cannot have $x\psi = y'\psi$, for $x\psi$ is a non-constant map and $y'\psi$ is a constant map. So to check injectivity, we simply check that $\psi|_M$ and $\psi_{M'}$ are injective:

$$x\psi = y\psi \Rightarrow \rho_x = \rho_y \Rightarrow 1\rho_x = 1\rho_y \Rightarrow x = y,$$
$$x'\psi = y'\psi \Rightarrow \tau_x = \tau_y \Rightarrow 1\tau_x = 1\tau_y \Rightarrow x = y.$$

Finally, to check that ψ is a homomorphism, we must check the various cases of multiplication in the definition of $C(M)$:

$$(x\varphi)(y'\varphi) = \rho_x\tau_y = \tau_y = y'\varphi = (xy')\varphi$$
$$(x'\varphi)(y'\varphi) = \tau_x\tau_y = \tau_y = y'\varphi = (x'y')\varphi$$
$$(x'\varphi)(y\varphi) = \tau_x\rho_y = \tau_{xy} = (xy)'\varphi.$$

So ψ is an isomorphism.

7.11 Let $x \in M$ and $y \in C(M)$. Then

$$\begin{aligned}
&(x)[(y)(f\,{}^{m}g)_{\mathrm{con}}] \\
&= ((x)f\,{}^{m}g)' && \text{[by definition of } {}_{\mathrm{con}}\text{]} \\
&= ((x)f)'(xm)g && \text{[by def. of the product and action]} \\
&= (x)[(y)f_{\mathrm{con}}](x)[(m')g_{\mathrm{ext}}] && \text{[by definition of } {}_{\mathrm{ext}} \text{ and } {}_{\mathrm{con}}\text{]} \\
&= (x)[(y)f_{\mathrm{con}}](x)[(ym')g_{\mathrm{ext}}] && \text{[by def. of the product in } C(M)\text{]} \\
&= (x)[(y)f_{\mathrm{con}}(y)\,{}^{m'}g_{\mathrm{ext}}] && \text{[by multiplication in } C(S)^M\text{]} \\
&= (x)[(y)f_{\mathrm{con}}\,{}^{m'}g_{\mathrm{ext}}]; && \text{[by multiplication in } (C(S)^M)^{C(M)}\text{]}
\end{aligned}$$

this proves (7.6). Next,

$$\begin{aligned}
&(x)[(y)g_{\mathrm{con}}] \\
&= ((x)g)' && \text{[by definition of } {}_{\mathrm{con}}\text{]} \\
&= (xy)f((x)g)' && \text{[by def. of the product in } C(S)\text{]} \\
&= (x)[(y)f_{\mathrm{ext}}](x)[(ym)g_{\mathrm{con}}] && \text{[by definition of } {}_{\mathrm{ext}} \text{ and } {}_{\mathrm{con}}\text{]} \\
&= (x)[(y)f_{\mathrm{ext}}(y)\,{}^{m}g_{\mathrm{con}}] && \text{[by multiplication in } C(S)^M\text{]} \\
&= (x)[(y)f_{\mathrm{ext}}\,{}^{m}g_{\mathrm{con}}]; && \text{[by multiplication in } (C(S)^M)^{C(M)}\text{]}
\end{aligned}$$

this proves (7.7). Finally,

$$\begin{aligned}
&(x)[(y)g_{\mathrm{con}}] \\
&= ((x)g)' && \text{[by definition of } {}_{\mathrm{con}}\text{]} \\
&= ((x)f)'((x)g)' && \text{[by def. of the product in } C(S)\text{]} \\
&= (x)[(y)f_{\mathrm{con}}](x)[(ym)g_{\mathrm{con}}] && \text{[by definition of } {}_{\mathrm{con}}\text{]} \\
&= (x)[(y)f_{\mathrm{con}}(y)\,{}^{m}g_{\mathrm{con}}] && \text{[by multiplication in } C(S)^M\text{]} \\
&= (x)[(y)f_{\mathrm{con}}\,{}^{m}g_{\mathrm{con}}]; && \text{[by multiplication in } (C(S)^M)^{C(M)}\text{]}
\end{aligned}$$

this proves (7.8).

EXERCISES FOR CHAPTER 8

[See pages 173–174 for the exercises.]

8.1 a) Suppose $w = u$. Then for any homomorphism $\vartheta : A^+ \to S$ we have $w\vartheta = u\vartheta = v\vartheta = (v'\vartheta)(w\vartheta)$. Then for any $q \in S$, we have $q(w\vartheta) = q(v'\vartheta)(w\vartheta)$ and so by cancellativity $q = q(v'\vartheta)$. So $v'\vartheta$ is a right identity for S and thus (by cancellativity) an identity. Let a and b be the first and last letters of v' (which may or may not be distinct). For $s \in S$, put $a\vartheta = b\vartheta = s$ to see that s is right and left invertible. Thus S is a group.

b) Suppose $w \neq u$. Since w is the longest common suffix of u and v, we know that u' and v' end with different letters a and b of A. That is, $u' = u''a$ and $v' = v''a$. Let $s, t \in S$. Let $\vartheta : A \to S$ be such that $a\vartheta = s$ and $b\vartheta = t$. Then $(u''\vartheta)s(w\vartheta) = u\vartheta = v\vartheta = (v''\vartheta)t(w\vartheta)$ and so $(u''\vartheta)s = (v''\vartheta)t$ by cancellativity. Hence s and t have a common left multiple. Since this holds for all $s, t \in S$, the semigroup S is group-embeddable by Exercise 6.5.

8.2 a) Let $\mathcal{N}$ be the class of finite nilpotent semigroups. Let $S \in \mathcal{N}$. So $S^n = \{0\}$ for some $n \in \mathbb{N}$. First, let T be a subsemigroup of S. Then $T^n \subseteq S^n = \{0\}$; hence $T \in \mathcal{N}$. So $\mathcal{N}$ is closed under $\mathbb{S}$. Second, let $\varphi : S \to U$ be a surjective homomorphism. Then $U^n = (S\varphi)^n = S^n\varphi \subseteq \{0_S\}\varphi = \{0_U\}$. So $U \in \mathcal{N}$. Thus $\mathcal{N}$ is closed under $\mathbb{H}$. Third, let $S_1, \ldots, S_k$ be nilpotent; then $S_i^{n_i} = \{0_{S_i}\}$ for some $n_i \in \mathbb{N}$ for each $i = 1, \ldots, k$. Let n be the maximum of the various n_i. Then

$$(S_1 \times \ldots \times S_k)^n \subseteq S_1^n \times \ldots S_k^n = \{0_{S_1}\} \times \ldots \{0_{S_k}\} = \{(0_{S_1}, \ldots, 0_{S_k})\};$$

hence $S_1 \times \ldots \times S_k \in \mathcal{N}$. Thus $\mathcal{N}$ is closed under $\mathbb{P}_{\text{fin}}$. Therefore $\mathcal{N}$ is a pseudovariety.

b) Let $A = \{a\}$. For each $k \in \mathbb{N}$, let $I_k = \{ w \in A^+ : |w| \geqslant k \}$. Then I_k is an ideal of A^+. Let $S_k = A^+/I_k$; then $S_k^k = \{0_{S_k}\}$. So each S_k is nilpotent. Let $S = \prod_{i=1}^{\infty} S_k$. Let $s \in S$ be such that $(k)s = a \in S_k$ for all $k \in \mathbb{N}$. Then for any $n \in \mathbb{N}$, we have $(n+1)s^n = a^n \in S_{n+1}$; hence $(n+1)s^n \neq 0_{S_{n+1}}$, and so $s^n \neq 0_S$ for any $n \in \mathbb{N}$. Thus $S^n \neq \{0_S\}$ for any $n \in \mathbb{N}$. Hence S is not nilpotent. Therefore the class of nilpotent semigroups is not closed under $\mathbb{P}$ and so is not a variety.

8.3 Note first that we are working with algebras of type $\{(\circ, 2), ({}^{-1}, 1)\}$. Let S be an orthodox completely regular semigroup. Let $\varphi : S \to T$ be a surjective homomorphism. Then T is regular by Proposition 4.20, and furthermore $(x\varphi)^{-1} = (x^{-1}\varphi)$ since homomorphisms for algebras of this type must also preserve ${}^{-1}$. Therefore since S is completely regular and thus satisfies the laws (4.2), T also satisfies these laws, so T is completely regular. Finally, if $e, f \in T$ are idempotents, then $e = xx^{-1}$ and $f = yy^{-1}$ for some $x, y \in T$ by Theorem 4.15. Let $p, q \in S$ be such that $p\varphi = x$ and $q\varphi = y$. Then pp^{-1} and qq^{-1} are idempotent. So $pp^{-1}qq^{-1}$ is idempotent (since S is orthodox) and so $(pp^{-1}qq^{-1})\varphi = xx^{-1}yy^{-1} = ef$ is idempotent. So the idempotents of T form a subsemigroup and so T is orthodox.

Now let T be a subalgebra of S. Then T also satisfies the laws (4.2) and is thus completely regular. Finally, the set of idempotents of T is the intersection of the set of idempotents of S, which is a subsemigroup, and T, which is also a subsemigroup. Hence the set of idempotents of T is a subsemigroup.

Finally, let $\{S_i : i \in I\}$ be a collection of orthodox completely regular semigroups. Then each S_i satisfies the laws (4.2) and so their product $\prod_{i\in I} S_i$ does also. The set of idempotents in $\prod_{i\in I} S_i$ is the product of the sets of idempotents in each S_i and hence forms a subsemigroup.

Now let S be an orthodox completely regular semigroup. Then S satisfies the laws (4.2). Let $x, y \in S$. Note that $x^{-1}x$ and yy^{-1} are idempotents, and so their product $x^{-1}xyy^{-1}$ is idempotent since S is orthodox. Thus

$$\begin{aligned} & xyy^{-1}x^{-1}xy \\ &= xx^{-1}xyy^{-1}x^{-1}xyy^{-1}y \\ &= xx^{-1}xyy^{-1}y && [\text{since } x^{-1}xyy^{-1} \text{ is idempotent}] \\ &= xy. \end{aligned}$$

Therefore S satisfies the law $xyy^{-1}x^{-1}xy = xy$.

Now suppose S satisfies the laws (4.2) and $xyy^{-1}x^{-1}xy = xy$. Then S is completely regular. Let $e, f \in S$ be idempotents; then $e = x^{-1}x$ and $f = yy^{-1}$ for some $x, y \in S$ by Theorem 4.15. Then

$$\begin{aligned} & (ef)^2 \\ &= (x^{-1}xyy^{-1})^2 \\ &= x^{-1}xyy^{-1}x^{-1}xyy^{-1} \\ &= x^{-1}xyy^{-1}; && [\text{since } xyy^{-1}x^{-1}xy = xy] \\ &= ef. \end{aligned}$$

Hence the idempotents of S form a subsemigroup and so S is orthodox.

8.4 a) Let $S = L \times R$ be a rectangular band, where L is a left zero semigroup and R is a right zero semigroup.

Let $\varphi : S \to T$ be a surjective homomorphism. Fix $(\ell, r) \in S$. Let $L_T = (L \times \{r\})\varphi$ and $R_T = (\{\ell\} \times R)\varphi$. Notice that L_T is a left zero semigroup and R_T is a right zero semigroup; hence $L_T \times R_T$ is a rectangular band. Define $\psi : L_T \times R_T \to T$ by $(\ell_t, r_t)\psi = \ell_t r_t$. Let $(\ell_t^{(1)}, r_t^{(1)}), (\ell_t^{(2)}, r_t^{(2)}) \in L_T \times R_T$. Let $\ell^{(1)}, \ell^{(2)} \in L$ and $r^{(1)}, r^{(2)} \in R$ be such that $(\ell^{(i)}, r)\varphi = \ell_t^{(i)}$ and $(\ell, r^{(i)})\varphi = r_t^{(i)}$ for $i = 1, 2$. Then

$$\begin{aligned} & (\ell_t^{(1)}, r_t^{(1)})\psi(\ell_t^{(2)}, r_t^{(2)})\psi \\ &= \ell_t^{(1)} r_t^{(1)} \ell_t^{(2)} r_t^{(2)} \\ &= (\ell^{(1)}, r)\varphi(\ell, r^{(1)})\varphi(\ell^{(2)}, r)\varphi(\ell, r^{(2)})\varphi \\ &= \big((\ell^{(1)}, r)(\ell, r^{(1)})(\ell^{(2)}, r)(\ell, r^{(2)})\big)\varphi \\ &= (\ell^{(1)}, r^{(2)})\varphi \\ &= \big((\ell^{(1)}, r)(\ell, r^{(2)})\big)\varphi \\ &= (\ell^{(1)}, r)\varphi(\ell, r^{(2)})\varphi \end{aligned}$$

$$
\begin{aligned}
&= \ell_t^{(1)} r_t^{(2)} \\
&= (\ell_t^{(1)}, r_t^{(2)})\psi \\
&= \big((\ell_t^{(1)}, r_t^{(1)})(\ell_t^{(2)}, r_t^{(2)})\big)\psi;
\end{aligned}
$$

thus ψ is a homomorphism. Furthermore,

$$
\begin{aligned}
&(\ell_t^{(1)}, r_t^{(1)})\psi = (\ell_t^{(2)}, r_t^{(2)})\psi \\
\Rightarrow\ & \ell_t^{(1)} r_t^{(1)} = \ell_t^{(2)} r_t^{(2)} \\
\Rightarrow\ & (\ell^{(1)}, r)\varphi(\ell, r^{(1)})\varphi = (\ell^{(2)}, r)\varphi(\ell, r^{(2)})\varphi \\
\Rightarrow\ & \big((\ell^{(1)}, r)(\ell, r^{(1)})\big)\varphi = \big((\ell^{(2)}, r)(\ell, r^{(2)})\big)\varphi \\
\Rightarrow\ & (\ell^{(1)}, r^{(1)})\varphi = (\ell^{(2)}, r^{(2)})\varphi \\
\Rightarrow\ & (\ell^{(1)}, r^{(1)})\varphi(\ell, r)\varphi = (\ell^{(2)}, r^{(2)})\varphi(\ell, r)\varphi \\
& \qquad \wedge (\ell, r)\varphi(\ell^{(1)}, r^{(1)})\varphi = (\ell, r)\varphi(\ell^{(2)}, r^{(2)})\varphi \\
\Rightarrow\ & \big((\ell^{(1)}, r^{(1)})(\ell, r)\big)\varphi = \big((\ell^{(2)}, r^{(2)})(\ell, r)\big)\varphi \\
& \qquad \wedge \big((\ell, r)(\ell^{(1)}, r^{(1)})\big)\varphi = \big((\ell, r)(\ell^{(2)}, r^{(2)})\big)\varphi \\
\Rightarrow\ & (\ell^{(1)}, r)\varphi = (\ell^{(2)}, r)\varphi \wedge (\ell, r^{(1)})\varphi = (\ell, r^{(2)})\varphi \\
\Rightarrow\ & \ell_t^{(1)} = \ell_t^{(2)} \wedge r_t^{(1)} = r_t^{(2)} \\
\Rightarrow\ & (\ell_t^{(1)}, r_t^{(1)}) = (\ell_t^{(2)}, r_t^{(2)}),
\end{aligned}
$$

so ψ is injective. Finally, ψ is surjective since

$$
\begin{aligned}
\operatorname{im}\psi &= L_T R_T \\
&= (L \times \{r\})\varphi(\{\ell\} \times R)\varphi \\
&= \big((L \times \{r\})(\{\ell\} \times R)\big)\varphi \\
&= (L \times R)\varphi = T.
\end{aligned}
$$

Hence T is isomorphic to the rectangular band $L_T \times R_T$; thus $T \in \mathsf{RB}$. So RB is closed under forming homomorphic images.

Now let T be a subsemigroup of S. Let $L_T = \{\, \ell \in L : (\exists r \in R)((\ell, r) \in T) \,\}$ and $R_T = \{\, r \in R : (\exists \ell \in L)((\ell, r) \in T) \,\}$. Notice that L_T is also a left zero semigroup and R_T is also a right zero semigroup. Clearly $T \subseteq L_T \times R_T$; we now establish the opposite inclusion. Let $(\ell_t, r_t) \in L_T \times R_T$. Then there exist $r \in R$ and $\ell \in L$ such that $(\ell_t, r) \in T$ and $(\ell, r_t) \in T$. Thus $(\ell_t, r_t) \in (\ell_t, r)(\ell, r_t) \in T$. So $T = L_T \times R_T$ is a rectangular band. So RB is closed under taking subsemigroups.

Finally, let $\{\, S_i : i \in I \,\}$ be a collection of rectangular bands. Then $S_i \simeq L_i \times R_i$ for some left zero semigroup L_i and right zero semigroup R_i, for each $i \in I$. Then

$$
\prod_{i \in I} S_i = \prod_{i \in I} (L_i \times R_i) \simeq \big(\prod_{i \in I} L_i\big) \times \big(\prod_{i \in I} R_i\big).
$$

Since $\prod_{i\in I} L_i$ is a left zero semigroup and $\prod_{i\in I} R_i$ is a right zero semigroup, $\prod_{i\in I} S_i \in \mathsf{RB}$. Hence RB is closed under forming direct products.

Thus RB is a variety.

b) Let $S = L\times R$ be a rectangular band. Let $x = (l_1, r_1)$ and $y = (l_2, r_2)$. Then $xyx = (l_1, r_1)(l_2, r_2)(l_1, r_1) = (l_1, r_1) = x$. So S satisfies this law.

Suppose S satisfies the law $xyx = x$. Fix some $t \in S$. Let $L = St$ and $R = tS$. Then for any $pt, p't \in L$, we have $ptp't = pt$ by the law (with $x = t$ and $y = p'$). So L is a left zero semigroup and similarly R is a right zero semigroup. Furthermore, for any $p, q.r \in S$,

$$\left.\begin{aligned} pr &= pqpr && \text{[by the law with } x = p \text{ and } y = q\text{]} \\ &= pqrqpr && \text{[by the law with } x = q \text{ and } y = r\text{]} \\ &= pqr. && \text{[by the law with } x = r \text{ and } y = qp\text{]} \end{aligned}\right\} \quad \text{(S.23)}$$

Define $\psi : S \to L \times R$ by $p\psi = (pt, tp)$. Then

$$\begin{aligned} (p\psi)(q\psi) &= (pt, tp)(qt, tq) \\ &= (pt, tq) \\ &= (pqt, tpq) && \text{[using (S.23) in both components]} \\ &= (pq)\psi, \end{aligned}$$

so ψ is a homomorphism. Notice that this also shows that for any $pt \in L$, $tq \in R$, we have $(pq)\psi = (pt, tq)$; thus ψ is surjective. Finally, for any $p, q \in S$,

$$\begin{aligned} & p\psi = q\psi \\ \Rightarrow\ & (pt, tp) = (qt, tq) \\ \Rightarrow\ & pt = qt \wedge tp = tq \\ \Rightarrow\ & ptp = qtp \wedge qtp = qtq \\ \Rightarrow\ & ptp = qtq \\ \Rightarrow\ & p = q, && \text{[applying the law on both sides]} \end{aligned}$$

so ψ is injective. So S is [isomorphic to] a rectangular band and so $S \in \mathsf{RB}$.

c) Any rectangular band satisfies the law $xyz = xz$ by (S.23). Every element of a rectangular band is idempotent, so $x^2 = x$ is also satisfied.

Let S satisfy the laws $x^2 = x$ and $xyz = xz$. To prove that S is a rectangular band, follow the reasoning in part b) with the following minor differences: First, L is a left zero semigroup since $ptp't = ptt = pz$ by applying first $xyz = xz$ and then $x^2 = x$;

similarly R is a right zero semigroup. Second, to prove ψ is a homomorphism, apply $xyz = xz$ to both components. Finally, the last step in proving ψ is injective becomes $ptp = qtq \Rightarrow p^2 = q^2 \Rightarrow p = q$, by applying first $xyz = xz$ and then $x^2 = x$.

d) Let S be a non-trivial null semigroup. Then for any $x, y, z \in S$, we have $xyz = 0_S$ and $xz = 0_S$. However, S is not a rectangular band because $x^2 \neq x$ for any $x \neq 0_S$.

8.5 Let $S = G \times L \times R$, where G is a group, L is a left zero semigroup, and R is a right zero semigroup. Let $\varphi : S \to T$ be a homomorphism. Fix $(1_G, \ell, r) \in S$. Let $H = (G \times \{\ell\} \times \{r\})\varphi$, $L_T = (\{1_G\} \times L \times \{r\})\varphi$ and $R_T = (\{1_G\} \times \{\ell\} \times R)\varphi$. Reasoning parallel to Example 8.4 shows that $T \simeq H \times L_T \times R_T$.

Notice that $(g, \ell, r)^{-1} = (g^{-1}, \ell, r)$. Let T be a subalgebra of S. Let $H = \{\, g \in G : (\exists (\ell, r) \in L \times R)((g, \ell, r) \in T \,\}$. We first prove that if $(g, \ell, r) \in T$, then $H \times \{(\ell, r)\} \subseteq T$. Let $h \in H$; then $(h, \ell', r') \in T$ for some $\ell' \in L$, $r' \in R$. Hence T contains

$$\begin{aligned} &(g, \ell, r)(g, \ell, r)^{-1}(h, \ell', r')(g, \ell, r)(g, \ell, r)^{-1} \\ &= (gg^{-1}hgg^{-1}, \ell, r) \\ &= (h, \ell, r), \end{aligned}$$

and thus $H \times \{(\ell, r)\} \subseteq T$. Now reason as in Example 8.4 to see that $T = H \times L_T \times R_T$ and thus $T \in \mathcal{X}$.

Let $\{\, S_i : i \in I \,\}$ be a collection of semigroups in $\mathcal{X}$. Then for all $i \in I$, we have $S_i \simeq G_i \times L_i \times R_i$ for some group G_i, left zero semigroup L_i and right zero semigroup R_i. Hence

$$\prod_{i \in I} S_i \simeq \prod_{i \in I} (G_i \times L_i \times R_i) \simeq \left(\prod_{i \in I} G_i\right) \times \left(\prod_{i \in I} L_i\right) \times \left(\prod_{i \in I} R_i\right);$$

since $\prod_{i \in I} G_i$ is a group, $\prod_{i \in I} L_i$ is a left zero semigroup, and $\prod_{i \in I} R_i$ is a right zero semigroup, we see that $\prod_{i \in I} S_i$ is [isomorphic to] the direct product of a group and a rectangular band. So $\prod_{i \in I} S_i \in \mathcal{X}$.

Let $S = G \times L \times R$, where G is a group, L is a left zero semigroup, and R is a right zero semigroup. Let $x = (g, \ell, r)$ and $y = (g', \ell', r')$. Then

$$\begin{aligned} xx^{-1} &= (g, \ell, r)(g^{-1}, \ell, r) \\ &= (1_G, \ell, r) \\ &= (g^{-1}, \ell, r)(g, \ell, r) \\ &= x^{-1}x \end{aligned}$$

and

$$\begin{aligned} x^{-1}yy^{-1}x &= (g^{-1}, \ell, r)(h, \ell', r')(h^{-1}, \ell', r')(g, \ell, r) \\ &= (g^{-1}hh^{-1}g, \ell, r) \end{aligned}$$

$$
\begin{aligned}
&= (1_G, \ell, r) \\
&= (g^{-1}, \ell, r)(g, \ell, r) \\
&= x^{-1}x.
\end{aligned}
$$

So S satisfies these laws.

Now suppose that S satisfies the given laws. For any $x, y \in S$, we have $x = xx^{-1}x = xx^{-1}yy^{-1}x \in SyS$. So S is simple by the analogue of Lemma 3.7 for simple semigroups. Let $e, f \in S$ be idempotents; then $e = xx^{-1}$ and $f = yy^{-1}$ for some $x, y \in S$. Then $efe = xx^{-1}yy^{-1}xx^{-1} = xx^{-1}xx^{-1} = xx^{-1} = e$. So the idempotents of S form a rectangular band by Example 8.4. Since rectangular bands are completely simple, they contain primitive idempotents. Hence S contains a primitive idempotent. So S is completely simple. Since the idempotents of S form a subsemigroup, S is orthodox. Hence S is a direct product of a rectangular band and a group by Exercise 5.6(b).

8.6 Let $S \in \bigcap_{i\in I} \mathsf{V}_i$. Then $S \in \mathsf{V}_i$ for all $i \in I$. Let T be a homomorphic image (respectively, subalgebra) of S. Since each V_i is a pseudovariety, $T \in \mathsf{V}_i$ for all $i \in I$. Hence $T \in \bigcap_{i\in I} \mathsf{V}_i$. So $\bigcap_{i\in I} \mathsf{V}_i$ is closed under forming homomorphic images and subalgebras. Now let $S_1, \dots, S_n \in \bigcap_{i\in I} \mathsf{V}_i$. Then $S_j \in \mathsf{V}_i$ for each $i \in I$ and $j = 1, \dots, n$. So $S_1 \times \ldots \times S_n \in \mathsf{V}_i$ for each $i \in I$ and so $S_1 \times \ldots \times S_n \in \bigcap_{i\in I} \mathsf{V}_i$. So $\bigcap_{i\in I} \mathsf{V}_i$ is closed under forming finitary direct products. Therefore $\bigcap_{i\in I} \mathsf{V}_i$ is a pseudovariety.

8.7 Let V be an $\mathcal{S}$-pseudovariety of semigroups. Then

$$
\begin{aligned}
& S \in (\mathsf{V}_{\mathrm{Mon}})_{\mathrm{Sg}} \\
\Rightarrow\ & S^1 \in \mathsf{V}_{\mathrm{Mon}} && \text{[by (8.7)]} \\
\Rightarrow\ & S^1 \text{ is a monoid in } \mathsf{V} \\
\Rightarrow\ & S \in \mathsf{V}. && \text{[since } S \text{ is closed under taking subsemigroups]}
\end{aligned}
$$

Let V be the $\mathcal{S}$-pseudovariety of rectangular bands. Then $\mathsf{V}_{\mathrm{Mon}} = \mathsf{1}$, since the only monoid that is a rectangular band is the trivial monoid, and so $(\mathsf{V}_{\mathrm{Mon}})_{\mathrm{Sg}} = \mathsf{V}_{\mathcal{S}}(1) = \mathsf{1} \neq \mathsf{V}$.

8.8 Let S be a completely regular semigroup. Let $s \in S$. By Theorem 4.15, s lies in a subgroup G of S. If $\vartheta : \bar{\Omega}_{\{x\}}\mathsf{S} \to S$ is such that $x\vartheta = s$, then $x^\omega\vartheta$ is the idempotent power of S, which much be the identity of G. So $x^{\omega+1}\vartheta = (x^\omega\vartheta)(x\vartheta) = 1s = s = x\vartheta$, so S satisfies the pseudoidentity $x^{\omega+1} = x$.

Now suppose that S satisfies $x^{\omega+1} = x$. Let $s \in S$ and choose $\vartheta : \bar{\Omega}_{\{x\}}\mathsf{S} \to S$ with $x\vartheta = s$. Then $x^\omega\vartheta = s^k$ for some $k \in \mathbb{N}$. So $s^k = x^\omega\vartheta = x\vartheta = s$. Thus s lies in the cyclic group $\{s, s^2, \dots, s^{k-1}\}$. Hence every element of S lies in a subgroup and so S is completely regular by Theorem 4.15.

8.9 Let S be a completely simple semigroup; thus $S = \mathcal{M}[G; I, \Lambda; P]$ for some group G, index sets I and Λ, and matrix P over G. Let (i, g, λ)

and (j,h,μ) be elements of S. If $\vartheta : \bar{\Omega}_{\{x\}}\mathsf{S} \to S$ is such that $x\vartheta = (i,g,\lambda)$ and $y\vartheta = (j,h,\mu)$, then we have $(xy)\vartheta = (i,g,\lambda)(j,h,\mu) = (i,gp_{\lambda j}h,\mu)$. Now, $(i,gp_{\lambda j}h,\mu)^k = (i,(gp_{\lambda j}hp_{\mu i})^{k-1}gp_{\lambda j}h,\mu)$ for all $k \in \mathbb{N}$. Thus $(xy)^\omega\vartheta$ is $(i,(gp_{\lambda j}hp_{\mu i})^{k-1}gp_{\lambda j}h,\mu)$ for some k. Since $(xy)^\omega\vartheta$ is always an idempotent, we have $(xy)^\omega\vartheta = (i,p_{\mu i}^{-1},\mu)$. Therefore we have $((xy)^\omega x)\vartheta = (i,p_{\mu i}^{-1},\mu)(i,g,h) = (i,p_{\mu i}^{-1},\mu)(i,g,\lambda) = (i,p_{\mu i}^{-1}p_{\mu i}g,\lambda) = (i,g,\lambda) = x\vartheta$. Thus S satisfies the pseudoidentity $(xy)^\omega x = x$.

Now suppose that S satisfies $(xy)^{\omega+1} = x$. Let $s,t \in S$ and choose $\vartheta : \bar{\Omega}_{\{x\}}\mathsf{S} \to S$ with $x\vartheta = s$ and $y\vartheta = t$. Then $(xy)^\omega x\vartheta = (st)^k s$ for some $k \in \mathbb{N}$. Hence $s = (st)^k s \in StS$ and so S is simple by the analogue of Lemma 3.7 for simple semigroups. Arguing as in Exercise 8.8 but with $x\vartheta = y\vartheta = s$, we see that s lies in the $\{s, s^2, \ldots, s^{2k}\}$. Hence every element of S lies in a subgroup and so S is completely regular by Theorem 4.15. Since S is completely regular and simple, it is completely simple by Theorem 4.16.

8.10 Let S be left simple. Let e be an idempotent of S. Then $Se = S$ since S is left simple. Let $s \in S$; then $s = s'e$ for some $s' \in S$. Therefore $se = s'ee = s'e = s$, and so e is a right identity for S. For any homomorphism $\vartheta : \bar{\Omega}_{\{x,y\}}(S)$, the element $y^\omega\vartheta$ is an idempotent of S. Hence $(xy^\omega)\vartheta = (x\vartheta)(y^\omega\vartheta) = x\vartheta$. Thus S satisfies the pseudoidentity $xy^\omega = x$.

Now suppose S satisfies $xy^\omega = x$. Let $s,t \in S$. Let $\vartheta : \bar{\Omega}_{\{x,y\}}\mathsf{S}$ be such that $x\vartheta = s$ and $y\vartheta = t$. Then $(y^\omega)\vartheta$ will be some idempotent power of t, say t^k for some $k \in \mathbb{N}$. Then $st^k = (xy^\omega)\vartheta = x\vartheta = s$. Hence $s \in St$. Thus $S = St$ for all $t \in S$ and so S is left simple.

EXERCISES FOR CHAPTER 9

[See pages 201–202 for the exercises.]

9.1 Suppose L is rational. Then it is recognized by a finite semigroup S by Theorem 9.4. By Proposition 9.6, SynM L divides S. Hence SynM L is finite.

Suppose SynM L is finite. The monoid SynM L recognizes L by Proposition 9.6. Since L is recognized by a finite monoid, it is rational by Theorem 9.4.

9.2 Let S be the three element semilattice $\{0,x,y\}$ with $x > 0$ and $y > 0$. Let $a\varphi = x$ and $b\varphi = y$. Then $\{x\}\varphi^{-1} = \{a\}^+$ and $\{y\}\varphi^{-1} = \{b\}^+$; hence $\{0\}\psi^{-1} = L$.

9.3 By definition, SynM $D = \{\,(\,,\,)\,\}^*/\sigma_D$. Let $w_1 \cdots w_n \in \{\,(\,,\,)\,\}^*$. Then

for any i, j,

$$C(w_1 \cdots w_j \,(\,)\, w_{j+1} \cdots w_n, i) = \begin{cases} C(w_1 \cdots w_n, i) & \text{if } i \leqslant j, \\ C(w_1 \cdots w_n, j) + 1 & \text{if } i = j+1, \\ C(w_1 \cdots w_n, j) & \text{if } i = j+2, \\ C(w_1 \cdots w_n, i) & \text{if } i \geqslant j+2. \end{cases}$$

In particular,

$$\begin{aligned} &C(w_1 \cdots w_j \,(\,)\, w_{j+1} \cdots w_n, n+2) = 0 \Leftrightarrow C(w_1 \cdots w_n, n) = 0, \\ &C(w_1 \cdots w_j \,(\,)\, w_{j+1} \cdots w_n, i) \geqslant 0 \text{ for all } i \\ &\qquad \Leftrightarrow C(w_1 \cdots w_n, i) \geqslant 0 \text{ for all } i. \end{aligned}$$

Hence for any words $p, q \in \{\,(\,,\,)\,\}^*$, we have $p\,(\,)\,q \in D$ if and only if $pq \in D$. Hence $(\,)\ \sigma_D\ \varepsilon$. That is, $[\,(\,]_{\sigma_D}[\,)\,]_{\sigma_D} = [\varepsilon]_{\sigma_D}$. Furthermore, $)\,($ is not a Dyck word, so $)\,($ is not σ_D-related to ε. That is $[\,)\,]_{\sigma_D}[\,(\,]_{\sigma_D} \neq [\varepsilon]_{\sigma_D}$. Hence, by Exercise 2.12 with $x = [\,(\,]_{\sigma_D}$, $y = [\,)\,]_{\sigma_D}$, and $e = [\varepsilon]_{\sigma_D}$, and noting that $[\,(\,]_{\sigma_D}$ and $y = [\,)\,]_{\sigma_D}$ generate SynM D, we see that SynM D is isomorphic to the bicyclic monoid.

9.4 Let $K, L \in \mathcal{N}(A^+)$. If both K and L are finite, then $K \cup L$ and $K \cap L$ are finite and so $K \cup L, K \cap L \in \mathcal{N}(A^+)$. If one of K or L is finite and the other cofinite, then $K \cup L$ is cofinite and $K \cap L$ is finite and so $K \cup L, K \cap L \in \mathcal{N}(A^+)$. If both K and L are cofinite, then $K \cup L$ and $K \cap L$ are cofinite and so $K \cup L, K \cap L \in \mathcal{N}(A^+)$. So $\mathcal{N}(A^+)$ is closed under union and intersection. If K is finite, $A^+ \setminus K$ is cofinite and so $A^+ \setminus K \in \mathcal{N}(A^+)$; if K is cofinite, $A^+ \setminus K$ is finite and so $A^+ \setminus K \in \mathcal{N}(A^+)$. So $\mathcal{N}(A^+)$ is closed under complementation.

Let $L \in \mathcal{N}(A^+)$ and $a \in A$. If L is finite, it contains only word of length less than n for some fixed $n \in \mathbb{N}$. So $a^{-1}L$ and La^{-1} contain only words of length less than $n-1$. So $a^{-1}L$ and La^{-1} are finite and so $a^{-1}L, La^{-1} \in \mathcal{N}(A^+)$. On the other hand, if L is cofinite, it contains all words in A^+ of length greater than n for some fixed $n \in \mathbb{N}$. So $a^{-1}L$ and La^{-1} contain all words in A^+ of length greater than $n-1$. So $a^{-1}L$ and La^{-1} are cofinite and so $a^{-1}L, La^{-1} \in \mathcal{N}(A^+)$.

Let $L \in \mathcal{N}(B^+)$ and let $\varphi : A^+ \to B^+$ be a homomorphism. If L is finite, it contains only word of length less than n for some fixed $n \in \mathbb{N}$. Let $w \in A^+$ have length greater than n. Then $w\varphi$ has length greater than n and so $w\varphi \notin L$. So $L\varphi^{-1}$ contains only words of length less than n; thus $L\varphi^{-1}$ is finite and so $L\varphi^{-1} \in \mathcal{N}(A^+)$. On the other hand, if L is cofinite, it contains all words in A^+ of length greater than n for some fixed $n \in \mathbb{N}$. Let $w \in A^+$ have length greater than n. Then $w\varphi$ has length greater than n and so $w\varphi \in L$. So $L\varphi^{-1}$ contains all words in A^+ of length greater than n; thus $L\varphi^{-1}$ is cofinite and so $L\varphi^{-1} \in \mathcal{N}(A^+)$.

9.5 Suppose that K is a $+$-language over A recognized by some finite rectangular band S. Then there is a homomorphism $\varphi : A^+ \to S$

such that $K = K\varphi\varphi^{-1}$. Recall from Exercise 8.4(c) that S satisfies the pseudoidentities $x^2 = x$ and $xyz = xz$. Thus, for $a, a' \in A$ and $w \in A^*$, we have $(awa')\varphi \in K\varphi$ if and only if $(aa')\varphi \in K\varphi$, or equivalently $awa' \in K$ if and only if $aa' \in K$. Therefore

$$\begin{aligned} & aA^*a' \cap K \neq \varnothing \\ \Rightarrow\ & (\exists u \in A^*)(aua' \in K) \\ \Rightarrow\ & aa' \in K \\ \Rightarrow\ & (\forall w \in A^*)(awa' \in K) \\ \Rightarrow\ & aA^*a' \subseteq K. \end{aligned}$$

On the other hand, if $aA^*a' \subseteq K$, then obviously $aA^*a' \cap K \neq \varnothing$. Therefore:

$$aA^*a' \subseteq K \Leftrightarrow aA^*a' \cap K \neq \varnothing. \tag{S.24}$$

Reasoning similar to the above and also using $(aa)\varphi = a\varphi$ proves that

$$aA^*a \subseteq K \Leftrightarrow a \in K. \tag{S.25}$$

Let Z be the subset of A that lies in K and let $K_1 = Z \cup \bigcup_{a \in Z} aA^*a$. Then by (S.25), $K_1 \subseteq K$. Again by (S.25), K_1 must be precisely the words in K that start and end with the same letter. Let K_2 be the set of words in K that start and end with different letters. By (S.24), if there a word in K_2 that starts with a and ends with a', then all words in aA^*a' lie in K_2. There are only finitely many possible choices for a and a', so $K_2 = \bigcup_{i=1}^{n} a_i A^* a'_i$ for suitable a_i and a'_i. Hence $K = K_1 \cup K_2$ is a language of the form (9.12).

Now suppose that K has the form (9.12). Then whether a word in A^+ lies in K depends only on its first and last letters. Let $s, t \in A^+$. Then for any $p, q \in A^*$, the first letters of $pstsq$ and psq are either both the first letter of p, and thus equal, or (when $p = \varepsilon$) both the first letter of s, and thus equal. Similarly, the last letters of $pstsq$ and psq are equal. So $pstsq \in K$ if and only if $psq \in K$. Hence $sts\ \sigma_K\ s$, or $[s]_{\sigma_K}[t]_{\sigma_K}[s]_{\sigma_K} = [s]_{\sigma_K}$. Since $s, t \in A^+$ were arbitrary, this proves that $\operatorname{Syn}_S K$ satisfies the pseudoidentity $xyx = x$. Hence $\operatorname{Syn}_S K$ is a rectangular band and $\operatorname{Syn}_S K \in \mathsf{RB}$.

Bibliography

> ‘The bibliographical references at the end of this book do not make up a bibliography, they are only a legal device aimed at avoiding accusations of having omitted the names of persons from whom I took direct quotations.’
>
> — Umberto Eco,
> *Kant and the Platypus*, Introduction
> (trans. William Weaver).

ALMEIDA, J. *Finite Semigroups and Universal Algebra*. Series in Algebra 3. World Scientific, 1994. ISBN: 9789810218959.

——— ‘Profinite semigroups and applications’. In: *Structural Theory of Automata, Semigroups, and Universal Algebra*. NATO Science Series II: Mathematics, Physics and Chemistry 207. Dordrecht: Springer, 2005, pp. 1–45. ISBN: 9781402038150. DOI: 10.1007/1-4020-3817-8_1.

ANDERSEN, O. ‘Ein bericht über die Struktur abstrakter Halbgruppen’. Staatsexamensarbeit. Hamburg, 1952.

BAADER, F. & NIPKOW, T. *Term Rewriting and All That*. Cambridge University Press, 1999. ISBN: 9780521455206.

BOOK, R. V. & OTTO, F. *String Rewriting Systems*. Texts and Monographs in Computer Science. Springer, 1993. ISBN: 9780387979656.

CAIN, A. J., ROBERTSON, E. F. & RUŠKUC, N. ‘Cancellative and Malcev presentations for finite Rees index subsemigroups and extensions’. In: *Journal of the Australian Mathematical Society* 84, no. 1 (2008), pp. 39–61. DOI: 10.1017/s1446788708000086.

CLIFFORD, A. H. & PRESTON, G. B. *The Algebraic Theory of Semigroups*. Vol. 1. Mathematical Surveys 7. Providence, RI: American Mathematical Society, 1961.

——— *The Algebraic Theory of Semigroups*. Vol. 2. Mathematical Surveys and Monographs 7. Providence, RI: American Mathematical Society, 1967.

CLIFFORD, A. H. ‘Semigroups admitting relative inverses’. In: *Annals of Mathematics*. 2nd ser. 42 (1941), pp. 1037–1049. DOI: 10.2307/1968781.

DISTLER, A. ‘Classification and Enumeration of Finite Semigroups’. Ph.D. thesis. University of St. Andrews, 2010. URL: http://hdl.handle.net/10023/945.

EILENBERG, S. *Automata, Languages, and Machines*. Vol. B. Pure and Applied Mathematics 59. New York: Academic Press, 1976.

GALLAGHER, P. 'On the Finite Generation and Presentability of Diagonal Acts, Finitary Power Semigroups and Schützenberger Products'. Ph.D. thesis. University of St Andrews, 2005.

GREEN, J. A. 'On the structure of semigroups'. In: *Annals of Mathematics*. 2nd ser. 54 (1951), pp. 163–172. DOI: 10.2307/1969317.

GRILLET, P. A. *Semigroups: An Introduction to the Structure Theory*. Monographs and Textbooks in Pure and Applied Mathematics 193. New York: Marcel Dekker, 1995. ISBN: 9780824796624.

——— *Commutative Semigroups*. Advances in Mathematics 2. Springer, 2011. ISBN: 9781441948571.

GRILLET, P. A. 'A short proof of Rédei's theorem'. In: *Semigroup Forum* 46, no. 1 (Dec. 1993), pp. 126–127. DOI: 10.1007/BF02573555.

HARJU, T. 'Lecture Notes on Semigroups'. Unpublished lecture notes, University of Turku. 1996. URL: http://users.utu.fi/harju/semigroups/semigroups96.pdf.

HIGGINS, P. M. *Techniques of Semigroup Theory*. With a forew. by G. B. Preston. Oxford Science Publications. Oxford: Clarendon Press, 1992. ISBN: 9780198535775.

HOPCROFT, J. E. & ULLMAN, J. D. *Introduction to Automata Theory, Languages, and Computation*. 1st ed. Reading, MA: Addison–Wesley, 1979. ISBN: 9780201029888.

HOWIE, J. M. *Automata and Languages*. Oxford Science Publications. Oxford: Clarendon Press, 1991. ISBN: 9780198534242.

——— *Fundamentals of Semigroup Theory*. London Mathematical Society Monographs: New Series 12. Oxford: Clarendon Press, 1995. ISBN: 9780198511946.

KLEENE, S. C. 'Representation of events in nerve nets and finite automata'. In: *Automata Studies*. Ed. by C. E. Shannon & J. McCarthy. Annals of Mathematics Studies 34. Princeton, NJ: Princeton University Press, 1956, pp. 3–41. ISBN: 978-0-691-07916-5.

KROHN, K. & RHODES, J. 'Algebraic theory of machines I: Prime decomposition theorem for finite semigroups and machines'. In: *Transactions of the American Mathematical Society* 116 (1965), pp. 450–464. DOI: 10.2307/1994127.

LALLEMENT, G. *Semigroups and Combinatorial Applications*. New York: John Wiley & Sons, 1979. ISBN: 9780471043799.

——— 'Augmentations and wreath products of monoids'. In: *Semigroup Forum* 21, no. 1 (Dec. 1980), pp. 89–90. DOI: 10.1007/BF02572539.

LAWSON, M. V. *Inverse Semigroups: The Theory of Partial Symmetries*. River Edge, NJ: World Scientific, 1998. ISBN: 9789810233167. DOI: 10.1142/9789812816689.

——— *Finite Automata*. Boca Raton, FL: Chapman & Hall/CRC, 2004. ISBN: 9781584882558.

LINDERHOLM, C. E. 'A group epimorphism is surjective'. In: *The American Mathematical Monthly* 77, no. 2 (Feb. 1970), p. 176. URL: https://www.jstor.org/stable/2317336.

LJAPIN, E. S. *Semigroups*. Trans. by A. A. Brown, J. M. Danskin, D. Foley, S. H. Gould, E. Hewitt, S. A. Walker & J. A. Zilber. 3rd ed. Translations of Mathematical Monographs 3. Providence, RI: American Mathematical Society, 1974.

LOTHAIRE, M. *Combinatorics on Words*. With a forew. by R. Lyndon. Corrected edition. Encyclopedia of Mathematics and its Applications 17. Cambridge University Press, 1997. ISBN: 9780521599245.

MAC LANE, S. *Categories for the Working Mathematician*. 2nd ed. Graduate Texts in Mathematics 5. Springer, 1998. ISBN: 9780387984032.

MALCEV, A. I. 'On the immersion of an algebraic ring into a field'. In: *Mathematische Annalen* 113 (1937), pp. 686–691. DOI: 10.1007/BF01571659.

MILLER, D. D. & CLIFFORD, A. H. 'Regular $\mathcal{D}$-classes in semigroups'. In: *Transactions of the American Mathematical Society* 82 (1956), pp. 270–280. DOI: 10.2307/1992989.

MUNN, W. D. 'Free Inverse Semigroups'. In: *Proceedings of the London Mathematical Society* 29, no. 3 (1st Nov. 1974), pp. 385–404. DOI: 10.1112/plms/s3-29.3.385.

ORE, Ø. 'Linear equations in non-commutative fields'. In: *Annals of Mathematics*. 2nd ser. 32, no. 3 (July 1931), pp. 463–477. DOI: 10.2307/1968245.

PETRICH, M. *Inverse Semigroups*. New York: John Wiley & Sons, 1984. ISBN: 9780471875451.

——— *Completely Regular Semigroups*. Canadian Mathematical Society Series of Monographs and Advanced Texts 23. New York: John Wiley & Sons, 1999. ISBN: 9780471195719.

PIN, J. E. *Varieties of Formal Languages*. Trans. by A. Howie. With a forew. by M. P. Schützenberger. Foundations of Computer Science. New York: Plenum Publishing, 1986. ISBN: 9781461293002. DOI: 10.1007/978-1-4613-2215-3.

PIN, J. 'Mathematical Foundations of Automata Theory'. Unpublished draft. 13th Mar. 2019.

PIN, J. 'Syntactic semigroups'. In: *Handbook of Formal Languages*. Vol. 1: *Word, Language, Grammar*. Ed. by G. Rozenberg & A. Salomaa. Berlin: Springer, 1997, pp. 679–746. ISBN: 9783642638633. DOI: 10.1007/978-3-642-59136-5_10.

PRESTON, G. B. 'Inverse semi-groups with minimal right ideals'. In: *Journal of the London Mathematical Society*. 1st ser. 29 (1954), pp. 404–411. DOI: 10.1112/jlms/s1-29.4.404.

——— 'Representations of inverse semi-groups'. In: *Journal of the London Mathematical Society*. 1st ser. 29 (1954), pp. 411–419. DOI: 10.1112/jlms/s1-29.4.411.

RABIN, M. O. & SCOTT, D. 'Finite automata and their decision problems'. In: *International Business Machines Journal of Research and Development* 3 (1959), pp. 114–125. DOI: 10.1147/rd.32.0114.

RÉDEI, L. *The Theory of Finitely Generated Commutative Semigroups*. Ed. by N. Reilly. International Series of Monographs in Pure and Applied Mathematics. Oxford: Pergamon Press, 1965.

REES, D. & HALL, P. 'On semi-groups'. In: *Mathematical Proceedings of the Cambridge Philosophical Society* 36, no. 04 (Oct. 1940), p. 387. DOI: 10.1017/S0305004100017436.

REES, D. 'On the group of a set of partial transformations'. In: *Journal of the London Mathematical Society*. 1st ser. 22, no. 4 (1947), pp. 281–284. DOI: 10.1112/jlms/s1-22.4.281.

RHODES, J. & STEINBERG, B. *The q-theory of Finite Semigroups*. Springer Monographs in Mathematics. New York: Springer, 2009. ISBN: 9780387097800. DOI: 10.1007/b104443.

ROBINSON, D. J. S. *A Course in the Theory of Groups*. Graduate Texts in Mathematics 80. Springer, 1995. ISBN: 9780387944616. DOI: 10.1007/978-1-4419-8594-1.

ROSALES, J. C. & GARCÍA-SÁNCHEZ, P. A. *Finitely Generated Commutative Monoids*. Commack, NY: Nova Science Publishers, 1999. ISBN: 9781560726708.

RUŠKUC, N. 'Semigroup Presentations'. Ph.D. thesis. University of St Andrews, 1995. URL: https://hdl.handle.net/10023/2821.

SCHÜTZENBERGER, M. P. 'On finite monoids having only trivial subgroups'. In: *Information and Control* 8, no. 2 (Apr. 1965), pp. 190–194. DOI: 10.1016/S0019-9958(65)90108-7.

SCHÜTZENBERGER, M.-P. '$\overline{D}$ représentation des demi-groupes'. In: *Comptes rendus hebdomadaires des séances de l'Académie des Sciences* 244, no. 2 (Apr.–June 1957), pp. 1994–1996. URL: https://gallica.bnf.fr/ark:/12148/bpt6k7215/f154.image.

——— 'Sur la représentation monomiale des demi-groupes'. In: *Comptes rendus hebdomadaires des séances de l'Académie des Sciences* 246, no. 1 (Jan.–Mar. 1958), pp. 865–867. URL: https://gallica.bnf.fr/ark:/12148/bpt6k3198s/f871.image.

SUSCHKEWITSCH, A. 'Über die endlichen Gruppen ohne das Gesetz der eindeutigen Umkehrbarkeit'. In: *Mathematische Annalen* 99 (1928), pp. 30–51. DOI: 10.1007/BF01459084.

VAGNER, V. V. 'Generalized groups'. In: *Doklady Akademii Nauk SSSR* 84 (1952), pp. 1119–1122.

Index

‘ Never index your own book. ’

— Kurt Vonnegut, *Cat's Cradle*, ch. 55.

In this index, the ordering of entries is strictly lexicographic, ignoring punctuation and spacing. Symbols outside the Latin alphabet are collected at the start of the index, even if ‘auxiliary’ Latin symbols are used: thus S^1 is included in this set, since it is the notation ‘ ^{1} ’ that is being defined. Brief definitions are given for notation.

This index currently covers only Chapter 1 and part of Chapter 3, plus names and ‘named results’. It will gradually be expanded to a full index.

$\wedge$: logical conjunction; ‘and’
0-simple, *58–60*
$\vee$: logical disjunction; ‘or’
1, 1_S: identity of a semigroup S
0, 0_S: zero of a semigroup S
S^1: monoid obtained by adjoining an identity to S if necessary; *4*
S^0: semigroup obtained by adjoining a zero to S if necessary; *4*
$\langle X \rangle$: subsemigroup generated by X; *10*
$\bigsqcap Y$, $x \sqcap y$: meet; *17*
$\bigsqcup Y$, $x \sqcup y$: join; *17*
ρ^{R}: reflexive closure of ρ; *22*
ρ^{S}: symmetric closure of ρ; *22*
ρ^{T}: transitive closure of ρ; *22*
ρ^{E}: equivalence relation generated by ρ; *22*
ρ^{C}: smallest left and right compatible relation containing ρ; *24*
$\rho^{\#}$: congruence generated by ρ; *24*
ρ_x: transformation that right-multiplies by x; *19*

action: *see* ‘semigroup action’
Almeida, Jorge, v, 174, 251
Andersen, Olaf, 71, 251
antichain, *15*
anti-homomorphism, *20*, 30
anti-symmetric binary relation, 15
associativity: *see* ‘binary operation, associative’
automaton: *see* ‘finite state automaton’; vi–vii

Baader, Franz, 53, 251
biber, 261
BibLaTeX, 261
bijection, 13
binary operation, *1*
  associative, 1–2
binary relation, *11–15*, 20, 22
Birkhoff's theorem, *154*
Book, Ronald V., 53, 251
Bourbaki, Nicolas, v
bracket, 2
Brown, Arthur A., 253
$\mathcal{B}_X$: set of binary relations on X; 12

Cain, Alan James, i–ii, vi–vii, 92, 251
cancellative semigroup, *6*, 7, 20, 32
  finite implies group, 32
cartesian product, *4*
  finitary, *4*
category theory, 1, 33, 35
Cayley graph, *30–31*
  of a group, 31
  right/left, 30–31
Cayley's theorem, 19
chain, *15*, 58
Chesterton, Gilbert Keith, 73
Clifford, A. H., 34, 70–71, 91–92, 119, 129, 251

Clifford, Alfred Hoblitzelle, 71, 92, 119, 251, 253
commutative semigroup, v–vi, *6*, *7–8*
  of idempotents, 18
comparable elements, 15
compatible binary relation, *20*, *24–26*
complete lattice: *see* 'lattice, complete'
complete lower semilattice: *see* 'semilattice, complete'
complete upper semilattice: *see* 'semilattice, complete'
composition of binary relations, 11
congruence, *20–21*, 28
  lattice of congruences, *26*
congruence generated by a binary relation, *24–26*
  characterization of, 25
converse of a relation, 11
correction, vi
coset, vii
Costa, Alfredo Manuel Gouveia da, 251
Cottingham, John, 175
Couto, Miguel Ângelo Marques Lourenço do, vii
Creative Commons, ii
cyclic group, 1

$\mathcal{D}$: *see also* 'Green's relation'; 55–57
D_a: $\mathcal{D}$-class of a; *57*
Danskin, John Moffatt, 253
Descartes, René, 175
determinant: *see* 'matrix, determinant of a'
dihedral group, 1
direct product, 1, *4*, 8, 28, 30
D'Israeli, Isaac, v
Distler, Andreas, 34, 174, 251
distributivity, 33
dom ρ: domain of ρ; *12*
Dyck, Walther Franz Anton von: *see* 'Dyck word'
Dyck word, *201–202*

$E(S)$: set of idempotents of S; *5*
Eco, Umberto, 251
Egri-Nagy, Attila, vii
Eilenberg correspondence, *189–201*
Eilenberg, Samuel: *see also* 'Eilenberg correspondence', 'Eilenberg's theorem'; 174, 202, 252
Eilenberg's theorem, *190*
'empty semigroup', 1, 35
End(S): endomorphisms of S; *19*
endomorphism, *19*
epimorphism, 34
  categorical, *33–34*
  of groups, 35
equivalence class, *15*, 20
equivalence relation, *15*, 22–27, 55
  characterization of join, 27
  commuting
    characterization of join, 27
  generated by a binary relation
    characterization of, 22
  lattice of equivalence relations, *26–27*
exponent, *5*
  laws, 5

factor group, vii
factor semigroup, *20–22*, 59–60
Feyeraband, Paul Karl, 149
finitely generated, *10*
finite semigroup, v–vi, 5, 33
  cancellative implies group, 32
finite state automaton, v
Foley, D., 253
'folklore', 34
free semigroup, vi
full map: *see* 'map'
full transformation: *see* 'transformation'

Gallagher, Peter Timothy, 53, 252
García Martinez, Xabier, vii
García-Sánchez, Pedro A., 129, 254
Gell-Mann, Murray, 121
generating set, *10*
Gould, Sydney Henry, 253
graph, vii
greatest lower bound: *see* 'meet'
Green, James Alexander: *see also* 'Green's lemma', 'Green's relations'; 71, 252
Green's lemma, 60
Green's relation: *see also* '$\mathcal{H}$, $\mathcal{L}$, $\mathcal{R}$, $\mathcal{D}$, $\mathcal{J}$'; *55–57*
  inclusion of, 56–57
  partial order from $\mathcal{L}$, $\mathcal{R}$, $\mathcal{J}$, 57
Grillet, Pierre Antoine, 34–35, 71, 129, 252
Grinberg, Darij, vii
group, vii, 1–2, *6*, 10, 32, 55–56, 58, 60
  composition series, 60
group-embeddable semigroup, *19*
group of units, *9*
groupoid, 1

$\mathcal{H}$: *see also* 'Green's relation'; 55, *56*
H_a: $\mathcal{H}$-class of a; *57*
Hall, P., 92, 254
Hamming, Richard Wesley, 203

Ham, Nick, vii
Hardy, Godfrey Harold, 1
Harju, Tero Juhani, 53, 252
Hasse diagram, 15–16, 18, 56
Herman, Samuel, vii
Hewitt, Edwin, 253
Higgins, Peter Michael, 34, 53, 252
homomorphic image, *19*
homomorphism, *19*–*20*, 21, 28–30, 33–34
  kernel of a, *19*, 21
  monoid, 19, 33
Hopcroft, John Edward, 202, 252
Howie, A., 253
Howie, John Mackintosh, v, 34, 53, 70, 91, 119, 174, 202, 252
Huxley, Thomas Henry, 131

$I(x)$: the set $J(x) \setminus J_x$; *59*
id_X: identity relation on X; *11*
ideal, *9*–*10*, 34, 55–60
  left: *see* 'left ideal'
  minimal
    uniqueness of, 58
  principal, *9*
  right: *see* 'right ideal'
  two-sided: *see* 'ideal'
ideal extension, *22*, 28–29
idempotent, *5*, 7–8, 32
  partial order of, *17*
idempotents
  semigroup of: *see* 'semigroup of idempotents'
identity: *see also* 'monoid'; 1, 3, 7, 9, 12, 32
  adjoining, *4*, 32
  left: *see* 'left identity'
  right: *see* 'right identity'
  two-sided: *see* 'identity'
  uniqueness of, 3
identity relation, *11*, 12, 19, 32
im ρ: image of ρ; *12*
index of an element, *5*
infimum: *see* 'meet'
integers
  as a partially ordered set, 15
  as a semigroup, 3
inverse, 1, *6*–*7*, 8
inverse semigroup, v–vi
invertible element, 6, 8–9, 33
isomorphism, *19*, 21

$\mathcal{J}$: *see also* 'Green's relation'; *55*, 56–57
J_a: $\mathcal{J}$-class of a; *57*
$J(x)$: principal ideal generated by x; *9*
join, *17*, 56
join semilattice: *see* 'semilattice'
'Jordan–Hölder theorem' for semigroups, 60

$K(S)$: kernel of a semigroup; *58*
$\ker \varphi$: kernel of the homomorphism φ; *19*
kernel, *58*, 59
  of a homomorphism: *see* 'homomorphism, kernel of a'
Kleene, Stephen Cole: *see also* 'Kleene's theorem'; 202, 252
Kleene's theorem, *180*
Knuth, Donald Ervin, v
Koga, Akihiko (古賀 明彦), vii
Korzybski, Alfred Habdank Skarbek, 55
Krohn, Kenneth Bruce: *see also* 'Krohn–Rhodes theorem'; 148, 252
Krohn–Rhodes theorem, *146*

$\mathcal{L}$: *see also* 'Green's relation'; *55*, 57
  commutes with $\mathcal{R}$, 56
L_a: $\mathcal{L}$-class of a; *57*
$L(x)$: principal left ideal generated by x; *9*
Lallement, Gérard, 148, 252
language, vii
  regular: *see* 'regular language'
lattice, *17*, 33
  complete, *17*
  of congruences: *see* 'congruence, lattice of congruences'
  of equivalence relation: *see also* 'equivalence relation, lattice of equivalence relations'
Lawson, Mark Verus, 119, 202, 252–253
least upper bound: *see* 'join'
left-cancellative semigroup, *6*, 32
left-compatible binary relation, *20*
left congruence, *20*, 57
left ideal, *9*–*10*
  0-minimal, *58*
  minimal, *58*
  principal, *9*
left identity, 3, 32
left inverse, *6*
left-invertible element, *6*, 33
left zero, 3, 32
left zero semigroup, *3*, 6, 8, 34
Linderholm, C. E., 35, 253
linear algebra, vii
Lisbon, i
Ljapin, Evgenii Sergeevich (Ляпин, Евгений Сергеевич), 34, 253
Lothaire, M., 53, 253
lower bound, *17*

lower semilattice: *see* 'semilattice'
LuaLaTeX, vi, 261
Lyndon, Roger Conant, 253

Mac Lane, Saunders, 35, 253
magma, 1
Malcev, Anatoly Ivanovich (Мальцев, Анатолий Иванович), 53, 253
Maltcev, Victor, vii
map, *12*
  domain of, *12*
  image of, *12*
  notation for, 4, 30
  preimage under, *12*
matrix
  determinant of a, 7
matrix semigroup, 7
maximal element, *16*
maximum element, *16*
McCarthy, John, 252
meet, *17*, 56
meet semilattice: *see* 'semilattice'
Miller, Don Dalzell, 71, 253
minimal element, *16*
minimum element, *16*
Mon$\langle X \rangle$: submonoid generated by X; *11*
monogenic semigroup, *10*
monoid: *see also* 'identity'; *3*, 7, 11–12, 28–29, 32–33
  presentation of: *see* 'monoid presentation'
  trivial: *see* 'trivial semigroup'
monomorphism, *19*, 34
  categorical, *33–34*
multiplication, *2*
Munn, William Douglas, 119, 253

natural homomorphism: *see* 'natural map'
natural map, *21*, 29
natural numbers
  as a semigroup, 2, 10, 30, 60
nilpotent group, 6
nilpotent semigroup, 5
nilsemigroup, 5
Nine Chapters on the Mathematical Art (九章算術; *Jiǔzhāng Suànshù*), vii
Nipkow, Tobias, 53, 251
null semigroup, *3*, 58–59

opposite semigroup, *8*
order, 15
Ore, Øystein: *see also* 'Ore's theorem'; 129, 253
Ore's theorem, *128*
Otto, Friedrich, 53, 251

$\mathbb{P}X$: power set; set of all subsets of X
$\mathcal{P}_X$: set of partial transformations on X; 12
partially ordered set, *15*, 16
  subset of, 15
partial map, *12*
partial order, *15–18*, 17
partial transformation, *12–13*
Pascal, Blaise, 93
periodic element, *5*
periodic semigroup, *5*, 33, 56–57
  infinite, 32
period of an element, *5*
Petrich, Mario, 92, 119, 253
PGF/Ti*k*Z, 261
Pin, Jean-Éric, v, 174, 202, 253
Porto, i
Porto, University of, vi
poset: *see* 'partially ordered set'
power, *5*
  positive, *5*
power semigroup, 32
power set, 15, 17
presentation, v–vi
  monoid: *see* 'monoid presentation'
  semigroup: *see* 'semigroup presentation'
Preston, Gordon Bamford: *see also* 'Vagner–Preston theorem'; 34, 70–71, 91–92, 119, 129, 251–253
principal factor, *59–60*
principal series, *60*
product of elements, *2*
product of subsets, *5*
pseudovariety, v–vi

quaternion group, 1
quotient semigroup: *see* 'factor semigroup'

$\mathcal{R}$: *see also* 'Green's relation'; *55*, 57
  commutes with $\mathcal{L}$, 56
R_a: $\mathcal{R}$-class of a; *57*
$R(x)$: principal right ideal generated by x; *9*
Rabin, Michael Oser (מִיכָאֵל עוזר רַבִּין), 202, 254
rectangular band, *8*
Rédei, László: *see also* 'Rédei's theorem'; 129, 254
Rédei's theorem, *127*
Rees congruence, *21*

Rees, David: *see also* 'Rees–Suschkewitsch theorem'; 92, 129, 254
Rees factor semigroup, *21–22*
Rees–Suschkewitsch theorem, *80*, *83*
reflexive binary relation, 15, 20, 22
reflexive closure of a binary relation, *22–24*
  characterization of, 22
regular element, *6–7*
regular language, v
regular semigroup, v–vi, *6*
Reilly, Norman R., 254
Reiterman's theorem, *167*
relation: *see* 'binary relation'
Rhodes, John Lewis: *see also* 'Krohn–Rhodes theorem'; 148, 174, 252, 254
right-cancellative semigroup, *6*
right-compatible binary relation, *20*
right congruence, *20*, 57
right ideal, *9–10*
  0-minimal, *58*
  minimal, *58*
  principal, *9*
right identity, 3, 32
right inverse, *6*
right-invertible element, *6*, 33
right regular representation, 19, 34
right zero, 3, 32
right zero semigroup, 3, 5, 8, 10, 31–32, 34
ring, 3
Rito, Guilherme Miguel Teixeira, vii
Robertson, Edmund Frederick, 92, 251
Robinson, Derek J. S., 71, 254
Rosales, José Carlos, 129, 254
Rozenberg, Grzegorz, 253
Ruškuc, Nikola, 53, 92, 251, 254

$\mathcal{S}_X$: set of bijections on X; 12
Salomaa, A., 253
Santiago de Compostella, University of, vi
Santos, José Manuel dos Santos dos, vii
Schelling, Friedrich Wilhelm Joseph von, 37
Schützenberger group, *65–68*
  right and left, *67*
Schützenberger, Marcel-Paul: *see also* 'Schützenberger group', 'Schützenberger's theorem'; 71, 202, 253–254
Schützenberger's theorem, *196*
Scott, Dana, 202, 254
semigroup, v–vi, *1–15*, 19–22, 24–35
  0-simple: *see* '0-simple semigroup'
  cancellative: *see* 'cancellative semigroup'
  commutative: *see* 'commutative semigroup'
  finite: *see* 'finite semigroup'
  free: *see* 'free semigroup'
  inverse: *see* 'inverse semigroup'
  left zero: *see* 'left zero semigroup'
  matrix: *see* 'matrix semigroup'
  null: *see* 'null semigroup'
  periodic: *see* 'periodic semigroup'
  presentation of: *see* 'semigroup presentation'
  regular: *see* 'regular semigroup'
  right zero: *see* 'right zero semigroup'
  simple: *see* 'simple semigroup'
  trivial: *see* 'trivial semigroup'
  zero-simple: *see* '0-simple semigroup'
semigroup action, *29–30*
  by endomorphisms, 30
  free, 30
  left, *30*
  regular, 30
  right, *30*
  transitive, 30
semigroup of binary relations, *12*, 13, 32
semigroup of idempotents, *5*
semigroup of partial transformations, 13, 30, 32
  computation(, 14
  computation), 14
semigroup of transformations, 13, 19, 29–30, 32–33
  computation(, 14
  computation), 14
semilattice, *17–18*, 34
  as a commutative semigroup of idempotents, 18
  complete, *17*
Shannon, Claude Elwood, 252
simple group, 58
simple semigroup, *58–60*
Soares, Jorge Fernando Valentim, vii
Steinberg, Benjamin, 148, 174, 254
structure of a semigroup, v–vi
subdirect product, *28–29*, 34
subgroup, *8–9*, 12, 19
submonoid: *see also* 'subsemigroup'; *8*, 12
  generating, *11*
subsemigroup: *see also* 'submonoid'; *8–10*, 19, 29, 32
  generating, 10
  proper, *8*, 32
supremum: *see* 'join'

Suschkewitsch, Anton Kazimirovich: *see also* 'Rees–Suschkewitsch theorem'
Suschkewitsch, Anton Kazimirovich (Сушкевич, Антон Казимирович), 92, 254
symmetric binary relation, *14–15*, 22
symmetric closure of a binary relation, *22–24*
  characterization of, 22
symmetric group, *12*, 13, 19, 30, 32

$\mathcal{T}_X$: set of transformations on X; 12
Tilson, Bret Ransom, 252
topology, vii
total order: *see* 'order'
transformation, *12–13*
  two-line notation, 13
transitive binary relation, 15, 20, 22
transitive closure of a binary relation, *22–24*
  characterization of, 22
trivial monoid: *see also* 'trivial semigroup'; *3*
trivial semigroup, *3*, 32
Trocado, Alexandre, vii
tuple, 4

Ullman, Jeffrey David, 202, 252
universal algebra, vii
upper bound, *17*
upper semilattice: *see* 'semilattice'

$V(x)$: set of inverses of x; *7*
Vagner–Preston theorem, *99*
Vagner, Viktor Vladimirovich: *see also* 'Vagner–Preston theorem'; 119, 254
variety: *see also* 'pseudovariety'; vi
Vonnegut, Kurt, 255

Walker, Sue Ann, 253
Weaver, William Fense, 251

`xindy`, 261

zero, 3, 6–7, 32, 58
  adjoining, *4*, 32
  left: *see* 'left zero'
  right: *see* 'right zero'
  two-sided: *see* 'zero'
  uniqueness of, 3
zero-simple semigroup: *see* '0-zero semigroup'
Zilber, Joseph Abraham, 253

COLOPHON

This document was typeset by the author using LuaLaTeX, with a custom style utilizing the packages `fontspec`, `unicode-math`, `microtype`, `titlesec`, `titletoc`, `booktabs`, `amsthm`, `amsmath`, and `mathtools`.

The main text is set in Minion Pro; captions and epigraphs are set in Myriad Pro. Mathematics is mainly set in Minion Math, with script letters from Minion Pro and sans-serif letters from Myriad Pro. Fixed-width text is set in `Noto Sans Mono`. Cyrillic text is set in Minion Pro. Chinese, Hebrew, and Japanese text is set in Noto.

Figures and commutative diagrams were created using PGF/Ti*k*Z.

The bibliography and citations were compiled using the BibLaTeX package and `biber` backend.

The index was compiled using `xindy` with a custom style.

