Mathematical Foundations of Automata Theory

Jean-Éric Pin

Version of March 24, 2025

Preface

These notes form the core of a future book on the algebraic foundations of automata theory. This book is still incomplete, but the first eleven chapters now form a relatively coherent material, covering roughly the topics described below.

The early years of automata theory

Kleene's theorem [69] is usually considered as the starting point of automata theory. It shows that the class of recognisable languages (that is, recognised by finite automata), coincides with the class of rational languages, which are given by rational expressions. Rational expressions can be thought of as a generalisation of polynomials involving three operations: union (which plays the role of addition), product and the star operation. It was quickly observed that these essentially combinatorial definitions can be interpreted in a very rich way in algebraic and logical terms. Automata over infinite words were introduced by Büchi in the early 1960's to solve decidability questions in first-order and monadic second-order logic of one successor. Investigating two-successor logic, Rabin was led to the concept of tree automata, which soon became a standard tool for studying logical definability.

The algebraic approach

The definition of the *syntactic monoid*, a monoid canonically attached to each language, was first given by Schützenberger in 1956 [140]. It later appeared in a paper of Rabin and Scott [131], where the notion is credited to Myhill. It was shown in particular that a language is recognisable if and only if its syntactic monoid is finite. However, the first classification results on recognisable languages were rather stated in terms of automata [91] and the first nontrivial use of the syntactic monoid is due to Schützenberger [141]. Schützenberger's theorem (1965) states that a rational language is star-free if and only if its syntactic monoid is finite and aperiodic. This elegant result is considered, right after Kleene's theorem, as the most important result of the algebraic theory of automata. Schützenberger's theorem was supplemented a few years later by a result of McNaughton [86], which establishes a link between star-free languages and first-order logic of the order relation.

Both results had a considerable influence on the theory. Two other important algebraic characterisations date back to the early seventies: Simon [145] proved that a rational language is piecewise testable if and only if its syntactic monoid is $\mathcal{J}$-trivial and Brzozowski-Simon [23] and independently, McNaughton [85]

characterised the locally testable languages. The logical counterpart of the first result was obtained by Thomas [168]. These successes settled the power of the algebraic approach, which was axiomatized by Eilenberg in 1976 [42].

Eilenberg's variety theory

A variety of finite monoids is a class of monoids closed under taking submonoids, quotients and finite direct products. Eilenberg's theorem states that varieties of finite monoids are in one-to-one correspondence with certain classes of recognisable languages, the varieties of languages. For instance, the rational languages are associated with the variety of all finite monoids, the star-free languages with the variety of finite aperiodic monoids, and the piecewise testable languages with the variety of finite $\mathcal{J}$-trivial monoids. Numerous similar results have been established over the past thirty years and, for this reason, the theory of finite automata is now intimately related to the theory of finite monoids.

Several attempts were made to extend Eilenberg's variety theory to a larger scope. For instance, partial order on syntactic semigroups were introduced in [101], leading to the notion of ordered syntactic semigroups. The resulting extension of Eilenberg's variety theory permits one to treat classes of languages that are not necessarily closed under complement, contrary to the original theory. Other extensions were developed independently by Straubing [162] and Ésik and Ito [45].

The topological point of view

Due allowance being made, the introduction of topology in automata theory can be compared to the use of p-adic analysis in number theory.

The notion of a variety of finite monoids was coined after a similar notion, introduced much earlier by Birkhoff for infinite monoids: a *Birkhoff variety of monoids* is a class of monoids closed under taking submonoids, quotient monoids and direct products. Birkhoff proved in [15] that his varieties can be defined by a set of identities: for instance the identity $xy = yx$ characterises the variety of commutative monoids. Almost fifty years later, Reiterman [133] extended Birkhoff's theorem to varieties of finite monoids: any variety of finite monoids can be characterised by a set of *profinite identities*. A profinite identity is an identity between two profinite words. Profinite words can be viewed as limits of sequences of words for a certain metric, the profinite metric. For instance, one can show that the sequence $x^{n!}$ converges to a profinite word denoted by x^ω and the variety of finite aperiodic monoids can be defined by the identity $x^\omega = x^{\omega+1}$.

The profinite approach is not only a powerful tool for studying varieties but it also led to unexpected developments, which are at the heart of the current research in this domain. In particular, Gehrke, Grigorieff and the author [47] proved that any lattice of recognisable languages can be defined by a set of profinite equations, a result that subsumes Eilenberg's variety theorem.

The logical approach

We already mentioned Büchi's, Rabin's and McNaughton's remarkable results on the connection between logic and finite automata. Büchi's *sequential calculus*

is a logical language to express combinatorial properties of words in a natural way. For instance, properties like "a word contains two consecutive occurrences of a" or "a word of even length" can be expressed in this logic. However, several parameters can be adjusted. Different fragments of logic can be considered: first-order, monadic second-order, Σ_n-formulas and a large variety of logical and nonlogical symbols can be employed.

There is a remarkable connection between first-order logic and the concatenation product. The polynomial closure of a class of languages $\mathcal{L}$ is the set of languages that are sums of marked products of languages of $\mathcal{L}$. By alternating *Boolean closure* and *polynomial closure*, one obtains a natural hierarchy of languages. The level 0 is the Boolean algebra $\{\emptyset, A^*\}$. Next, for each $n \geqslant 0$, the level $2n+1$ is the polynomial closure of the level $2n$ and the level $2n+2$ is the Boolean closure of the level $2n+1$. A very nice result of Thomas [168] shows that a recognisable language is of level $2n+1$ in this hierarchy if and only if it is definable by a Σ_{n+1}-sentence of first-order logic in the signature $\{<, (\mathbf{a})_{a\in A}\}$, where $\mathbf{a}$ is a predicate giving the positions of the letter a.

There are known algebraic characterisations for the three first levels of this hierarchy. In particular, the second level is the class of piecewise testable languages characterised by Simon [144].

Contents of these notes

The algebraic approach to automata theory relies mostly on semigroup theory, a branch of algebra which is usually not part of the standard background of a student in mathematics or in computer science. For this reason, an important part of these notes is devoted to an introduction to semigroup theory. Chapter II gives the basic definitions and Chapter V presents the structure theory of finite semigroups. Chapters XVI and XVIII introduce some more advanced tools, the relational morphisms and the semidirect and wreath products.

Chapter III gives a brief overview on finite automata and recognisable languages. It contains in particular a complete proof of Kleene's theorem which relies on Glushkov's algorithm in one direction and on linear equations in the opposite direction. For a comprehensive presentation of this theory I recommend the book of my colleague Jacques Sakarovitch [138]. The recent book of Olivier Carton [27] also contains a nice presentation of the basic properties of finite automata. Recognisable and rational subsets of a monoid are presented in Chapter IV. The notion of a syntactic monoid is the key notion of this chapter, where we also discuss the ordered case. Chapters VI and VII present two major results, at the core of the algebraic approach to automata theory: Schützenberger's and Simon's theorem. The profinite topology is introduced in Chapter X. We start with a short synopsis on general topology and metric spaces and then discuss the relationship between profinite topology and recognisable languages. Chapter XI is devoted to varieties of finite monoids and to Reiterman's theorem. It also contains a large collection of examples. Chapter XII presents the equational characterisation of lattices of languages. Eilenberg's variety theory forms the topic of Chapter XIII. Examples of application of these two results are gathered in Chapter XIV. The last five chapters are still under construction. Chapter XV is about polynomial closure, Chapter XVII presents another deep result of Schützenberger about unambiguous star-free languages and its logical counterpart. Chapter XIX gives a brief introduction to sequential

functions and the wreath product principle. Chapter IX presents some logical descriptions of languages and their algebraic characterisations.

Notation and terminology

The term *regular set* is frequently used in the literature but there is some confusion on its interpretation. In Ginzburg [51] and in Hopcroft, Motwani and Ullman [63], a regular set is a set of words accepted by a finite automaton. In Salomaa [139], it is a set of words defined by a regular grammar and in Carroll and Long [26], it is a set defined by a regular expression. This is no real problem for languages, since, by Kleene's theorem, these three definitions are equivalent. This is more problematic for monoids in which Kleene's theorem does not hold. Another source of confusion is that the term *regular* has a well-established meaning in semigroup theory. For these reasons, I prefer to use the terms *recognisable* and *rational*.

I tried to keep some homogeneity in notation. Most of the time, I use Greek letters for functions, lower case letters for elements, capital letters for sets and calligraphic letters for sets of sets. Thus I write: "let s be an element of a semigroup S and let $\mathcal{P}(S)$ be the set of subsets of S". I write functions on the left and transformations and actions on the right. In particular, I denote by $q\cdot u$ the action of a word u on a state q. Why so many computer scientists prefer the awful notation $\delta(q, u)$ is still a mystery. It leads to heavy formulas, like $\delta(\delta(q, u), v) = \delta(q, uv)$, to be compared to the simple and intuitive $(q\cdot u)\cdot v = q\cdot uv$, for absolutely no benefit.

I followed Eilenberg's tradition to use boldface letters, like $\mathbf{V}$, to denote varieties of semigroups, and to use calligraphic letters, like $\mathcal{V}$, for varieties of languages. However, I have adopted Almeida's suggestion to have a different notation for operators on varieties, like $\mathbb{E}\mathbf{V}$, $\mathbb{L}\mathbf{V}$ or $\mathbb{P}\mathbf{V}$.

I use the term *morphism* for *homomorphism*. Semigroups are usually denoted by S or T, monoids by M or N, alphabets are A or B and letters by a, b, c, ... but this notation is not frozen: I may also use A for semigroup and S for alphabet if needed! Following a tradition in combinatorics, $|E|$ denotes the number of elements of a finite set. The notation $|u|$ is also used for the length of a word u, but in practice, there is no risk of confusion between the two.

To avoid repetitions, I frequently use brackets as an equivalent to "respectively", like in the following sentence : a semigroup [monoid, group] S is commutative if, for all $x, y \in S$, $xy = yx$.

Lemmas, propositions, theorems and corollaries share the same counter and are numbered by section. Examples have a separate counter, but are also numbered by section. References are given according to the following example: Theorem 1.6, Corollary 1.5 and Section 1.2 refer to statements or sections of the same chapter. Proposition X.3.16 refers to a proposition which is external to the current chapter.

Acknowledgements

Several books on semigroups helped me in preparing these notes. Clifford and Preston's treatise [30, 31] remains the classical reference. My favourite source for the structure theory is Grillet's remarkable presentation [55]. I also borrowed a lot from the books by Almeida [4], Eilenberg [42], Higgins [60], Lallement [78]

and Lothaire [81] and also of course from my own books [98, 93]. Another source of inspiration (not yet fully explored!) are the research articles by my colleagues Jorge Almeida, Karl Auinger, Jean-Camille Birget, Olivier Carton, Mai Gehrke, Victor Guba, Rostislav Horčík, John McCammond, Stuart W. Margolis, Dominique Perrin, Mark Sapir, Imre Simon, Ben Steinberg, Howard Straubing, Pascal Tesson, Denis Thérien, Misha Volkov, Pascal Weil and Marc Zeitoun.

I would like to thank my former Ph.D. students Laure Daviaud, Luc Dartois, Charles Paperman and Yann Pequignot, my colleagues at IRIF and LaBRI and my former students of the Master Parisien de Recherches en Informatique (notably Aiswarya Cyriac, Nathanaël Fijalkow, Agnes Köhler, Arthur Milchior, Anca Nitulescu, Pierre Pradic, Léo Stefanesco, Amrita Suresh, Boker Udi, Jill-Jênn Vie and Furcy) for pointing out many misprints and corrections on earlier versions of this document. I would like to acknowledge the assistance and the encouragements of my colleagues of the Picasso project, Adolfo Ballester-Bolinches, Antonio Cano Gómez, Ramon Esteban-Romero, Xaro Soler-Escrivà, Maria Belén Soler Monreal, Jorge Palanca and of the Pessoa project, Jorge Almeida, Mário J. J. Branco, Vítor Hugo Fernandes, Gracinda M. S. Gomes and Pedro V. Silva. Other careful readers include Achim Blumensath and Martin Beaudry (with the help of his student Cédric Pinard) who proceeded to a very careful reading of the manuscript. George Hansoul, Alfonso Labao, Sébastien Labbé, Nathan Lothe, Amaldev Manuel, Anne Schilling, Manon Stipulanti and Herbert Toth sent me some very useful remarks. Special thanks are due to Jean Berstel and to Paul Gastin for providing me with their providential LaTeX packages.

Paris, March 2025

Jean-Éric Pin

Contents

Part A

Automata and semigroups

Chapter I

Algebraic preliminaries

1 Subsets, relations and functions

1.1 Sets

The set of subsets of a set E is denoted by $\mathcal{P}(E)$ (or sometimes 2^E). The *positive Boolean operations* on $\mathcal{P}(E)$ comprise *union* and *intersection*. The *Boolean operations* also include *complementation*. The complement of a subset X of E is denoted by X^c. Thus, for all subsets X and Y of E, the following relations hold

$$(X^c)^c = X \qquad (X \cup Y)^c = X^c \cap Y^c \qquad (X \cap Y)^c = X^c \cup Y^c$$

We let $|E|$ denote the number of elements of a finite set E, also called the *size* of E. A *singleton* is a set of size 1. We shall frequently identify a singleton $\{s\}$ with its unique element s.

Given two sets E and F, the set of ordered pairs (x, y) such that $x \in E$ and $y \in F$ is written $E \times F$ and called the *product* of E and F.

1.2 Relations

Let E and F be two sets. A *relation* on E and F is a subset of $E \times F$. If $E = F$, it is simply called a *relation on* E. A relation τ can also be viewed as a function[1] from E to $\mathcal{P}(F)$ by setting, for each $x \in E$,

$$\tau(x) = \{y \in F \mid (x, y) \in \tau\}$$

By abuse of language, we say that τ is a relation from E to F.

The *inverse of a relation* $\tau \subseteq E \times F$ is the relation $\tau^{-1} \subseteq F \times E$ defined by

$$\tau^{-1} = \{(y, x) \in F \times E \mid (x, y) \in \tau\}$$

Note that if τ is a relation from E to F, the relation τ^{-1} can be also viewed as a function from F to $\mathcal{P}(E)$ defined by

$$\tau^{-1}(y) = \{x \in E \mid y \in \tau(x)\}$$

[1] Functions are formally defined in the next section, but we assume the reader is already familiar with this notion.

A relation from E to F can be extended to a function from $\mathcal{P}(E)$ to $\mathcal{P}(F)$ by setting, for each subset X of E,

$$\tau(X) = \bigcup_{x \in X} \tau(x) = \{y \in F \mid \text{for some } x \in X,\ (x,y) \in \tau\}$$

If Y is a subset of F, we then have

$$\begin{aligned}\tau^{-1}(Y) &= \bigcup_{y \in Y} \tau^{-1}(y) = \{x \in E \mid \text{there exists } y \in Y \text{ such that } y \in \tau(x)\} \\ &= \{x \in E \mid \tau(x) \cap Y \neq \emptyset\}\end{aligned}$$

Given two relations $\tau_1 : E \to F$ and $\tau_2 : F \to G$, we let $\tau_1\tau_2$ or $\tau_2 \circ \tau_1$ denote the *composition* of τ_1 and τ_2, which is the relation from E to G defined by

$$(\tau_2 \circ \tau_1)(x) = \{z \in G \mid \text{there exists } y \in F \text{ such that } y \in \tau_1(x) \text{ and } z \in \tau_2(y)\}$$

1.3 Functions

A [partial] *function* $\varphi : E \to F$ is a relation on E and F such that for every $x \in E$, there exists one and only one [in the case of a partial function, at most one] element $y \in F$ such that $(x,y) \in \varphi$. When this y exists, it is denoted by $\varphi(x)$. The set

$$\mathrm{Dom}(\varphi) = \{x \in E \mid \text{there exists } y \in F \text{ such that } (x,y) \in \varphi\}$$

is called the *domain* of φ. A function with domain E is sometimes called a *total function* or a *mapping* from E to F. The set

$$\mathrm{Im}(\varphi) = \{y \in F \mid \text{there exists } x \in E \text{ such that } (x,y) \in \varphi\}$$

is called the *range* or the *image* of φ. Given a set E, the identity mapping on E is the mapping $\mathrm{Id}_E : E \to E$ defined by $\mathrm{Id}_E(x) = x$ for all $x \in E$.

A mapping $\varphi : E \to F$ is called *injective* if, for every $u, v \in E$, $\varphi(u) = \varphi(v)$ implies $u = v$. It is *surjective* if, for every $v \in F$, there exists $u \in E$ such that $v \in \varphi(u)$. It is *bijective* if it is simultaneously injective and surjective. For instance, the identity mapping $\mathrm{Id}_E(x)$ is bijective.

Proposition 1.1. *Let $\varphi : E \to F$ be a mapping. Then φ is surjective if and only if there exists a mapping $\psi : F \to E$ such that $\varphi \circ \psi = \mathrm{Id}_F$.*

Proof. If there exists a mapping ψ with these properties, we have $\varphi(\psi(y)) = y$ for all $y \in F$ and thus φ is surjective. Conversely, suppose that φ is surjective. For each element $y \in F$, select an element $\psi(y)$ in the nonempty set $\varphi^{-1}(y)$. This defines a mapping $\psi : F \to E$ such that $\varphi \circ \psi(y) = y$ for all $y \in F$. □

A consequence of Proposition 1.1 is that surjective maps are *right cancellative* (the definition of a right cancellative map is transparent, but if needed, a formal definition is given in Section II.1.2).

Corollary 1.2. *Let $\varphi : E \to F$ be a surjective mapping and let α and β be two mappings from F to G. If $\alpha \circ \varphi = \beta \circ \varphi$, then $\alpha = \beta$.*

Proof. By Proposition 1.1, there exists a mapping $\psi : F \to E$ such that $\varphi \circ \psi = \mathrm{Id}_F$. Therefore $\alpha \circ \varphi = \beta \circ \varphi$ implies $\alpha \circ \varphi \circ \psi = \beta \circ \varphi \circ \psi$, whence $\alpha = \beta$. □

Proposition 1.3. *Let $\varphi : E \to F$ be a mapping. Then φ is injective if and only if there exists a mapping $\psi : \mathrm{Im}(\varphi) \to E$ such that $\psi \circ \varphi = \mathrm{Id}_E$.*

Proof. Suppose there exists a mapping ψ with these properties. Then φ is injective since the condition $\varphi(x) = \varphi(y)$ implies $\psi \circ \varphi(x) = \psi \circ \varphi(y)$, that is, $x = y$. Conversely, suppose that φ is injective. Define a mapping $\psi : \mathrm{Im}(\varphi) \to E$ by setting $\psi(y) = x$, where x is the unique element of E such that $\varphi(x) = y$. Then $\psi \circ \varphi = \mathrm{Id}_E$ by construction. □

It follows that injective maps are *left cancellative*.

Corollary 1.4. *Let $\varphi : F \to G$ be an injective mapping and let α and β be two mappings from E to F. If $\varphi \circ \alpha = \varphi \circ \beta$, then $\alpha = \beta$.*

Proof. By Proposition 1.3, there exists a mapping $\psi : \mathrm{Im}(\varphi) \to F$ such that $\psi \circ \varphi = \mathrm{Id}_F$. Therefore $\varphi \circ \alpha = \varphi \circ \beta$ implies $\psi \circ \varphi \circ \alpha = \psi \circ \varphi \circ \beta$, whence $\alpha = \beta$. □

We come to a standard property of bijective maps.

Proposition 1.5. *If $\varphi : E \to F$ is a bijective mapping, there exists a unique bijective mapping from F to E, denoted by φ^{-1}, such that $\varphi \circ \varphi^{-1} = \mathrm{Id}_F$ and $\varphi^{-1} \circ \varphi = \mathrm{Id}_E$.*

Proof. Since φ is bijective, for each $y \in F$ there exists a unique $x \in E$ such that $\varphi(x) = y$. Thus the condition $\varphi^{-1} \circ \varphi = \mathrm{Id}_E$ requires that $x = \varphi^{-1}(\varphi(x)) = \varphi^{-1}(y)$. This ensures the uniqueness of the solution. Now, the mapping $\varphi^{-1} : F \to E$ defined by $\varphi^{-1}(y) = x$, where x is the unique element such that $\varphi(x) = y$, clearly satisfies the two conditions $\varphi \circ \varphi^{-1} = \mathrm{Id}_F$ and $\varphi^{-1} \circ \varphi = \mathrm{Id}_E$. □

The mapping φ^{-1} is called the *inverse* of φ.

It is clear that the composition of two injective [surjective] mappings is injective [surjective]. A partial converse to this result is given in the next proposition.

Proposition 1.6. *Let $\alpha : E \to F$ and $\beta : F \to G$ be two mappings and let $\gamma = \beta \circ \alpha$ be their composition.*

(1) *If γ is injective, then α is injective. If furthermore α is surjective, then β is injective.*

(2) *If γ is surjective, then β is surjective. If furthermore β is injective, then α is surjective.*

Proof. (1) Suppose that γ is injective. If $\alpha(x) = \alpha(y)$, then $\beta(\alpha(x)) = \beta(\alpha(y))$, whence $\gamma(x) = \gamma(y)$ and $x = y$ since γ is injective. Thus α is injective. If, furthermore, α is surjective, then it is bijective and, by Proposition 1.5, $\gamma \circ \alpha^{-1} = \beta \circ \alpha \circ \alpha^{-1} = \beta$. It follows that β is the composition of the two injective maps γ and α^{-1} and hence is injective.

(2) Suppose that γ is surjective. Then for each $z \in G$, there exists $x \in E$ such that $\gamma(x) = z$. It follows that $z = \beta(\alpha(x))$ and thus β is surjective. If, furthermore, β is injective, then it is bijective and, by Proposition 1.5, $\beta^{-1} \circ \gamma = \beta^{-1} \circ \beta \circ \alpha = \alpha$. It follows that α is the composition of the two surjective maps β^{-1} and γ and hence is surjective. □

The next result is extremely useful.

Proposition 1.7. *Let E and F be two finite sets such that $|E| = |F|$ and let $\varphi : E \to F$ be a function. The following conditions are equivalent:*

(1) *φ is injective,*

(2) *φ is surjective,*

(3) *φ is bijective.*

Proof. Clearly it suffices to show that (1) and (2) are equivalent. If φ is injective, then φ induces a bijection from E onto $\varphi(E)$. Thus $|E| = |\varphi(E)| \leqslant |F| = |E|$, whence $|\varphi(E)| = |F|$ and $\varphi(E) = F$ since F is finite.

Conversely, suppose that φ is surjective. By Proposition 1.1, there exists a mapping $\psi : F \to E$ such that $\varphi \circ \psi = \mathrm{Id}_F$. Since ψ is injective by Proposition 1.6, and since we have already proved that (1) implies (2), ψ is surjective. It follows by Proposition 1.6 that φ is injective. □

1.4 Injective and surjective relations

The notions of surjective and injective functions can be extended to relations as follows. A relation $\tau : E \to F$ is *surjective* if, for every $v \in F$, there exists $u \in E$ such that $v \in \tau(u)$. It is called *injective* if, for every $u, v \in E$, $\tau(u) \cap \tau(v) \neq \emptyset$ implies $u = v$. The next three propositions provide equivalent definitions.

Proposition 1.8. *A relation is injective if and only if its inverse is a partial function.*

Proof. Let $\tau : E \to F$ be a relation. Suppose that τ is injective. If $y_1, y_2 \in \tau^{-1}(x)$, then $x \in \tau(y_1) \cap \tau(y_2)$ and thus $y_1 = y_2$ since τ is injective. Thus τ^{-1} is a partial function.

Suppose now that τ^{-1} is a partial function. If $\tau(x) \cap \tau(y) \neq \emptyset$, there exists some element c in $\tau(x) \cap \tau(y)$. It follows that $x, y \in \tau^{-1}(c)$ and thus $x = y$ since τ^{-1} is a partial function. □

Proposition 1.9. *Let $\tau : E \to F$ be a relation. Then τ is injective if and only if, for all $X, Y \subseteq E$, $X \cap Y = \emptyset$ implies $\tau(X) \cap \tau(Y) = \emptyset$.*

Proof. Suppose that τ is injective and let X and Y be two disjoint subsets of E. If $\tau(X) \cap \tau(Y) \neq \emptyset$, then $\tau(x) \cap \tau(y) \neq \emptyset$ for some $x \in X$ and $y \in Y$. Since τ is injective, it follows that $x = y$ and hence $X \cap Y \neq \emptyset$, a contradiction. Thus $X \cap Y = \emptyset$.

If the condition of the statement holds, then it can be applied in particular when X and Y are singletons, say $X = \{x\}$ and $Y = \{y\}$. Then the condition becomes $x \neq y$ implies $\tau(x) \cap \tau(y) = \emptyset$, that is, τ is injective. □

Proposition 1.10. *Let $\tau : E \to F$ be a relation. The following conditions are equivalent:*

(1) *τ is injective,*

(2) *$\tau^{-1} \circ \tau = \mathrm{Id}_{\mathrm{Dom}(\tau)}$,*

(3) *$\tau^{-1} \circ \tau \subseteq \mathrm{Id}_E$.*

Proof. (1) implies (2). Suppose that τ is injective and let $y \in \tau^{-1} \circ \tau(x)$. By definition, there exists $z \in \tau(x)$ such that $y \in \tau^{-1}(z)$. Thus $x \in \mathrm{Dom}(\tau)$ and $z \in \tau(y)$. Now, $\tau(x) \cap \tau(y) \neq \emptyset$ and since τ is injective, $x = y$. Therefore $\tau^{-1} \circ \tau \subseteq \mathrm{Id}_{\mathrm{Dom}(\tau)}$. But if $x \in \mathrm{Dom}(\tau)$, there exists by definition $y \in \tau(x)$ and thus $x \in \tau^{-1} \circ \tau(x)$. Thus $\tau^{-1} \circ \tau = \mathrm{Id}_{\mathrm{Dom}(\tau)}$.
(2) implies (3) is trivial.
(3) implies (1). Suppose that $\tau^{-1} \circ \tau \subseteq \mathrm{Id}_E$ and let $x, y \in E$. If $\tau(x) \cap \tau(y)$ contains an element z, then $z \in \tau(x)$, $z \in \tau(y)$ and $y \in \tau^{-1}(z)$, whence $y \in \tau^{-1} \circ \tau(x)$. Since $\tau^{-1} \circ \tau \subseteq \mathrm{Id}_E$, it follows that $y = x$ and thus τ is injective. □

Proposition 1.10 has two useful consequences.

Corollary 1.11. *Let $\tau : E \to F$ be a relation. The following conditions are equivalent:*

(1) *τ is a partial function,*
(2) *$\tau \circ \tau^{-1} = \mathrm{Id}_{\mathrm{Im}(\tau)}$,*
(3) *$\tau \circ \tau^{-1} \subseteq \mathrm{Id}_F$.*

Proof. The result follows from Proposition 1.10 since, by Proposition 1.8, τ is a partial function if and only if τ^{-1} is injective. □

Corollary 1.12. *Let $\tau : E \to F$ be a relation. Then τ is a surjective partial function if and only if $\tau \circ \tau^{-1} = \mathrm{Id}_F$.*

Proof. Suppose that τ is a surjective partial function. Then by Corollary 1.11, $\tau \circ \tau^{-1} = \mathrm{Id}_F$.

Conversely, if $\tau \circ \tau^{-1} = \mathrm{Id}_F$, then τ is a partial function by Corollary 1.11 and $\tau \circ \tau^{-1} = \mathrm{Id}_{\mathrm{Im}(\tau)}$. Therefore $\mathrm{Im}(\tau) = F$ and τ is surjective. □

Corollary 1.12 is often used in the following form.

Corollary 1.13. *Let $\alpha : F \to G$ and $\beta : E \to F$ be two functions and let $\gamma = \alpha \circ \beta$. If β is surjective, the relation $\gamma \circ \beta^{-1}$ is equal to α.*

Proof. Indeed, $\gamma = \alpha \circ \beta$ implies $\gamma \circ \beta^{-1} = \alpha \circ \beta \circ \beta^{-1}$. Now, by Corollary 1.12, $\beta \circ \beta^{-1} = \mathrm{Id}_F$. Thus $\gamma \circ \beta^{-1} = \alpha$. □

It is clear that the composition of two injective [surjective] relations is injective [surjective]. Proposition 1.6 can also be partially adapted to relations.

Proposition 1.14. *Let $\alpha : E \to F$ and $\beta : F \to G$ be two relations and let $\gamma = \beta \circ \alpha$ be their composition.*

(1) *If γ is injective and β^{-1} is surjective, then α is injective. If furthermore α is a surjective partial function, then β is injective.*
(2) *If γ is surjective, then β is surjective. If furthermore β is injective of domain F, then α is surjective.*

Proof. (1) Suppose that γ is injective. If $\alpha(x) \cap \alpha(y) \neq \emptyset$, there exists an element $z \in \alpha(x) \cap \alpha(y)$. Since β^{-1} is surjective, there is a t such that $z \in \beta^{-1}(t)$ or, equivalently, $t \in \beta(z)$. Then $t \in \beta(\alpha(x)) \cap \beta(\alpha(y))$, whence $\gamma(x) = \gamma(y)$ and $x = y$ since γ is injective. Thus α is injective.

If furthermore α is a surjective partial function, then by Proposition 1.8, α^{-1} is an injective relation and by Corollary 1.12, $\alpha \circ \alpha^{-1} = \mathrm{Id}_F$. It follows

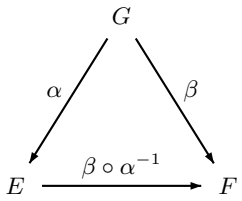
that $\gamma \circ \alpha^{-1} = \beta \circ \alpha \circ \alpha^{-1} = \beta$. Thus β is the composition of the two injective relations γ and α^{-1} and hence is injective.

(2) Suppose that γ is surjective. Then for each $z \in G$, there exists $x \in E$ such that $z \in \gamma(x)$. Thus there exists $y \in \alpha(x)$ such that $z \in \beta(y)$, which shows that β is surjective.

Suppose that β is injective of domain F or, equivalently, that β^{-1} is a surjective partial map. Then by Proposition 1.10, $\beta^{-1} \circ \beta = \mathrm{Id}_F$. It follows that $\beta^{-1} \circ \gamma = \beta^{-1} \circ \beta \circ \alpha = \alpha$. Therefore α is the composition of the two surjective relations β^{-1} and γ and hence is surjective. □

Proposition 1.15. *Let E, F, G be three sets and $\alpha : G \to E$ and $\beta : G \to F$ be two functions. Suppose that α is surjective and that, for every $s, t \in G$, $\alpha(s) = \alpha(t)$ implies $\beta(s) = \beta(t)$. Then the relation $\beta \circ \alpha^{-1} : E \to F$ is a function.*

Proof. Let $x \in E$. Since α is surjective, there exists $y \in G$ such that $\alpha(y) = x$. Setting $z = \beta(y)$, one has $z \in \beta \circ \alpha^{-1}(x)$.



Let $z' \in \beta \circ \alpha^{-1}(x)$. Then $z' = \beta(y')$ for some $y' \in \alpha^{-1}(x)$. Thus $\alpha(y') = x$ and since $\alpha(y) = \alpha(y')$, the condition of the statement implies that $\beta(y) = \beta(y')$. Thus $z = z'$, which shows that $\beta \circ \alpha^{-1}$ is a function. □

1.5 Relations and set operations

We gather in this section three elementary properties of relations. The first two propositions can be summarised by saying that "union commutes with relations", "Boolean operations commute with inverses of functions", "union, intersection and set difference commute with injective relations". The last one is a more subtle property of surjective partial functions.

Proposition 1.16. *Let $\tau : E \to F$ be a relation. Then for every $X, Y \subseteq E$, the relation $\tau(X \cup Y) = \tau(X) \cup \tau(Y)$ holds.*

Proof. It follows immediately from the definition:

$$\tau(X \cup Y) = \bigcup_{z \in X \cup Y} \tau(x) = \Big(\bigcup_{z \in X} \tau(x)\Big) \cup \Big(\bigcup_{z \in Y} \tau(x)\Big) = \tau(X) \cup \tau(Y). \qquad \square$$

Proposition 1.17. *Let $\tau : E \to F$ be an injective relation. Then, for every $X, Y \subseteq E$, the following relations hold:*

$$\tau(X \cup Y) = \tau(X) \cup \tau(Y) \quad \tau(X \cap Y) = \tau(X) \cap \tau(Y) \quad \tau(X - Y) = \tau(X) - \tau(Y).$$

Proof. The first formula follows from Proposition 1.16.

It follows from the inclusion $X \cap Y \subseteq X$ that $\tau(X \cap Y) \subseteq \tau(X)$ and similarly $\tau(X \cap Y) \subseteq \tau(Y)$. Thus $\tau(X \cap Y) \subseteq \tau(X) \cap \tau(Y)$. Now, if $z \in \tau(X) \cap \tau(Y)$, then $z \in \tau(x) \cap \tau(y)$ for some $x \in X$ and $y \in Y$. But since τ is injective, it follows $x = y$ and thus $z \in \tau(X \cap Y)$. Thus $\tau(X) \cap \tau(Y) \subseteq \tau(X \cap Y)$, which proves the second equation.

The first relation gives $\tau(X - Y) \cup \tau(Y) = \tau(X \cup Y)$. Thus

$$\tau(X) - \tau(Y) \subseteq \tau(X \cup Y) - \tau(Y) \subseteq \tau(X - Y)$$

Furthermore, $\tau(X - Y) \subseteq \tau(X)$ and since τ is injective, $\tau(X - Y) \cap \tau(Y) = \emptyset$ by Proposition 1.9. Thus $\tau(X - Y) \subseteq \tau(X) - \tau(Y)$ and finally $\tau(X - Y) = \tau(X) - \tau(Y)$, which proves the third relation. □

More precise results hold for inverses of functions on the one hand, and for surjective partial functions on the other hand.

Proposition 1.18. *Let $\varphi : E \to F$ be a function. Then, for every $X, Y \subseteq F$, the following relations hold:*

$$\begin{aligned} \varphi^{-1}(X \cup Y) &= \varphi^{-1}(X) \cup \varphi^{-1}(Y) \\ \varphi^{-1}(X \cap Y) &= \varphi^{-1}(X) \cap \varphi^{-1}(Y) \\ \varphi^{-1}(X^c) &= (\varphi^{-1}(X))^c. \end{aligned}$$

Proof. By Proposition 1.8, the relation φ^{-1} is injective and thus Proposition 1.17 gives the first two formulas. The third one relies on the fact that $\varphi^{-1}(F) = E$. Indeed, the third property of Proposition 1.17 gives $\varphi^{-1}(X^c) = \varphi^{-1}(F - X) = \varphi^{-1}(F) - \varphi^{-1}(X) = E - \varphi^{-1}(X) = (\varphi^{-1}(X))^c$. □

Proposition 1.19. *Let $\varphi : E \to F$ be a partial function. Then for every $X \subseteq E$ and $Y \subseteq F$, the following relation hold:*

$$\varphi(X) \cap Y = \varphi\big(X \cap \varphi^{-1}(Y)\big)$$

Proof. Clearly, $\varphi\big(X \cap \varphi^{-1}(Y)\big) \subseteq \varphi(X)$ and $\varphi\big(X \cap \varphi^{-1}(Y)\big) \subseteq \varphi(\varphi^{-1}(Y)) \subseteq Y$. Thus $\varphi\big(X \cap \varphi^{-1}(Y)\big) \subseteq \varphi(X) \cap Y$. Moreover, if $y \in \varphi(X) \cap Y$, then $y = \varphi(x)$ for some $x \in X$ and since $y \in Y$, $x \in \varphi^{-1}(Y)$. It follows that $\varphi(X) \cap Y \subseteq \varphi\big(X \cap \varphi^{-1}(Y)\big)$, which concludes the proof. □

2 Ordered sets

If $\mathcal{R}$ is a relation on E, two elements x and y of E are said to be *related by* $\mathcal{R}$ if $(x, y) \in \mathcal{R}$, which is also denoted by $x \mathcal{R} y$.

A relation $\mathcal{R}$ is *reflexive* if, for each $x \in E$, $x \mathcal{R} x$, *symmetric* if, for each $x, y \in E$, $x \mathcal{R} y$ implies $y \mathcal{R} x$, *antisymmetric* if, for each $x, y \in E$, $x \mathcal{R} y$ and $y \mathcal{R} x$ imply $x = y$ and *transitive* if, for each $x, y, z \in E$, $x \mathcal{R} y$ and $y \mathcal{R} z$ implies $x \mathcal{R} z$.

A relation is a *preorder* if it is reflexive and transitive, an *order* (or *partial order*) if it is reflexive, transitive and antisymmetric and an *equivalence relation* (or an *equivalence*) if it is reflexive, transitive and symmetric. If $\mathcal{R}$ is a preorder,

the relation $\sim$ defined by $x \sim y$ if and only if $x \mathcal{R} y$ and $y \mathcal{R} x$ is an equivalence relation, called the *equivalence relation associated* with $\mathcal{R}$.

Relations are ordered by inclusion. More precisely, if $\mathcal{R}_1$ and $\mathcal{R}_2$ are two relations on a set S, $\mathcal{R}_1$ *refines* $\mathcal{R}_2$ (or $\mathcal{R}_1$ is *thinner* than $\mathcal{R}_2$, or $\mathcal{R}_2$ is *coarser* than $\mathcal{R}_1$) if and only if, for each $s, t \in S$, $s \mathcal{R}_1 t$ implies $s \mathcal{R}_2 t$. Equality is thus the thinnest equivalence relation and the *universal* relation, in which all elements are related, is the coarsest. The following property is obvious.

Proposition 2.20. *Any intersection of preorders is a preorder. Any intersection of equivalence relations is an equivalence relation.*

It follows that, given a set R of relations on a set E, there is a least preorder [equivalence relation] containing all the relations of E. This relation is called the *preorder* [equivalence relation] *generated* by R.

Proposition 2.21. *Let $\mathcal{R}_1$ and $\mathcal{R}_2$ be two preorders [equivalence relations] on a set E. If they commute, the preorder [equivalence relation] generated by $\mathcal{R}_1$ and $\mathcal{R}_2$ is equal to $\mathcal{R}_1 \circ \mathcal{R}_2$.*

Proof. Suppose that $\mathcal{R}_1$ and $\mathcal{R}_2$ commute and let $\mathcal{R} = \mathcal{R}_1 \circ \mathcal{R}_2$. Then $\mathcal{R}$ is clearly reflexive but it is also transitive. Indeed, if $x_1 \mathcal{R} x_2$ and $x_2 \mathcal{R} x_3$, there exist $y_1, y_2 \in E$ such that $x_1 \mathcal{R}_1 y_1$, $y_1 \mathcal{R}_2 x_2$, $x_2 \mathcal{R}_1 y_2$ and $y_2 \mathcal{R}_2 x_3$. Since $\mathcal{R} = \mathcal{R}_2 \circ \mathcal{R}_1$, one gets $y_1 \mathcal{R} y_2$ and since $\mathcal{R}_1$ and $\mathcal{R}_2$ commute, there exists y such that $y_1 \mathcal{R}_1 y$ and $y \mathcal{R}_2 y_2$. It follows that $x_1 \mathcal{R}_1 y$ and $y \mathcal{R}_2 x_3$ and thus $x_1 \mathcal{R} x_3$. □

□

An *upper set* of an ordered set $(E, \leqslant)$ is a subset F of E such that, if $x \leqslant y$ and $x \in F$, then $y \in F$. The upper set generated by an element x is the set $\uparrow x$ of all $y \in E$ such that $x \leqslant y$. The intersection [union] of any family of upper sets is also an upper set.

A *chain* is a totally ordered subset of a partially ordered set.

3 Exercises

Exercise 1. Let $\varphi_1 : E_1 \to E_2$, $\varphi_2 : E_2 \to E_3$ and $\varphi_3 : E_3 \to E_1$ be three functions. Show that if, among the mappings $\varphi_3 \circ \varphi_2 \circ \varphi_1$, $\varphi_2 \circ \varphi_1 \circ \varphi_3$, and $\varphi_1 \circ \varphi_3 \circ \varphi_2$, two are surjective and the third is injective, or two are injective and the third is surjective, then the three functions φ_1, φ_2 and φ_3 are bijective.

Exercise 2. Let $\varphi : E \to F$ be a function. Show that φ is surjective if and only if, for every $T \subseteq F$, $\varphi(\varphi^{-1}(T)) = T$. Show that φ is injective if and only if, for every $S \subseteq E$, $\varphi^{-1}(\varphi(S)) = S$.

Exercise 3. This exercise is due to J. Almeida [4, Lemma 8.2.5].
Let X and Y be two totally ordered finite sets and let P be a partially ordered set. Let $\varphi : X \to Y$, $\gamma : Y \to X$, $\pi : X \to P$ and $\theta : Y \to P$ be functions such that:

(1) for any $x \in X$, $\pi(x) \leqslant \theta(\varphi(x))$,
(2) for any $y \in Y$, $\theta(y) \leqslant \pi(\gamma(y))$,
(3) if $x_1, x_2 \in X$, $\varphi(x_1) = \varphi(x_2)$ and $\pi(x_1) = \theta(\varphi(x_1))$, then $x_1 = x_2$,

(4) if $y_1, y_2 \in Y$, $\gamma(y_1) = \gamma(y_2)$ and $\theta(y_1) = \pi(\gamma(y_1))$, then $y_1 = y_2$.

Then φ and γ are mutually inverse functions, $\pi = \theta \circ \varphi$ and $\theta = \pi \circ \gamma$.

Chapter II

Semigroups and beyond

The purpose of this chapter is to introduce the basic algebraic definitions that will be used in this book: semigroups, monoids, groups and semirings, morphisms, substructures and quotients for these structures, transformation semigroups and free semigroups. We also devote a short section to idempotents, a central notion in finite semigroup theory.

1 Semigroups, monoids and groups

1.1 Semigroups, monoids

Let S be a set. A *binary operation* on S is a mapping from $S \times S$ to S. The image of (x, y) under this mapping is often denoted by xy and is called the *product* of x and y. In this case, it is convenient to call the binary operation *multiplication*. Sometimes, the additive notation $x+y$ is adopted, the operation is called *addition* and $x + y$ denotes the *sum* of x and y.

An operation on S is *associative* if, for every x, y, z in S, $(xy)z = x(yz)$. It is *commutative*, if, for every x, y in S, $xy = yx$.

An element 1 of S is called an *identity element* or simply *identity* for the operation if, for all $x \in S$, $x1 = x = 1x$. It is easy to see there can be at most one identity, which is then called *the* identity. Indeed if 1 and $1'$ are two identities, one has simultaneously $11' = 1'$, since 1 is an identity, and $11' = 1$, since $1'$ is an identity, whence $1 = 1'$. In additive notation, the identity is denoted by 0 to get the intuitive formula $0 + x = x = x + 0$.

A *semigroup* is a pair consisting of a set S and an associative binary operation on S. A semigroup is a pair, but we shall usually say "S is a semigroup" and assume the binary operation is known. A *monoid* is a triple consisting of a set M, an associative binary operation on M and an identity for this operation. The number of elements of a semigroup is often called its *order*[1]. Thus a semigroup of order 4 is a semigroup with 4 elements.

The *dual semigroup* of a semigroup S, denoted by $\tilde{S}$, is the semigroup defined on the set S by the operation $*$ given by $s * t = ts$.

A semigroup [monoid, group] is said to be *commutative* if its operation is commutative.

[1] This well-established terminology has of course nothing to do with the order relations.

If S is a semigroup, let S^1 denote the monoid equal to S if S is a monoid, and to $S \cup \{1\}$ if S is not a monoid. In the latter case, the operation of S is completed by the rules

$$1s = s1 = s$$

for each $s \in S^1$.

1.2 Special elements

Being idempotent, zero and cancellable are the three important properties of an element of a semigroup defined in this section. We also define the notion of semigroup inverse of an element. Regular elements, which form another important category of elements, will be introduced in Section V.2.2.

Idempotents

Let S be a semigroup. An element e of S is an *idempotent* if $e = e^2$. The set of idempotents of S is denoted by $E(S)$. We shall see later that idempotents play a fundamental role in the study of finite semigroups.

An element e of S is a *right identity* [*left identity*] of S if, for all $s \in S$, $se = s$ [$es = s$]. Observe that e is an identity if and only if it is simultaneously a right and a left identity. Furthermore, a right [left] identity is necessarily idempotent. The following elementary result illustrates these notions.

Proposition 1.1 (Simplification lemma)**.** *Let S be a semigroup. Let $s \in S$ and e, f be idempotents of S^1. If $s = esf$, then $es = s = sf$.*

Proof. If $s = esf$, then $es = eesf = esf = s$ and $sf = esff = esf = s$. □

Zeros

An element e is said to be a *zero* [*right zero*, *left zero*] if, for all $s \in S$, $es = e = se$ [$se = e$, $es = e$].

Proposition 1.2. *A semigroup has at most one zero element.*

Proof. Assume that e and e' are zero elements of a semigroup S. Then by definition, $e = ee' = e'$ and thus $e = e'$. □

If S is a semigroup, let S^0 denote the semigroup obtained from S by addition of a zero: the support[2] of S^0 is the disjoint union of S and the singleton[3] 0 and the multiplication (here denoted by $*$) is defined by

$$s * t = \begin{cases} st & \text{if } s, t \in S \\ 0 & \text{if } s = 0 \text{ or } t = 0. \end{cases}$$

A semigroup is called *null* if it has a zero and if the product of two elements is always zero.

[2] That is, the set on which the semigroup is defined.
[3] A singleton $\{s\}$ will also be denoted by s.

Cancellative elements

An element s of a semigroup S is said to be *right cancellative* [*left cancellative*] if, for every $x, y \in S$, the condition $xs = ys$ [$sx = sy$] implies $x = y$. It is *cancellative* if it is simultaneously right and left cancellative.

A semigroup S is *right cancellative* [*left cancellative*, *cancellative*] if all its elements are right cancellative [left cancellative, cancellative].

Inverses

We have to face a slight terminology problem with the notion of an inverse. Indeed, semigroup theorists have coined a notion of inverse that differs from the standard notion used in group theory and elsewhere in mathematics. Usually, the context should permit to clarify which definition is understood. But to avoid any ambiguity, we shall use the terms *group inverse* and *semigroup inverse* when we need to distinguish the two notions.

Let M be a monoid. A *right group inverse* [*left group inverse*] of an element x of M is an element x' such that $xx' = 1$ [$x'x = 1$]. A *group inverse* of x is an element x' which is simultaneously a right and left group inverse of x, so that $xx' = x'x = 1$.

We now come to the notion introduced by semigroup theorists. Given an element x of a semigroup S, an element x' is a *semigroup inverse* or simply *inverse* of x if $xx'x = x$ and $x'xx' = x'$.

It is clear that any group inverse is a semigroup inverse but the converse is not true. A thorough study of semigroup inverses will be given in Section V.2, but let us warn the reader immediately of some counterintuitive facts about inverses. An element of an infinite monoid may have several right group inverses and several left group inverses. The situation is radically different for a finite monoid: each element has at most one left [right] group inverse and if these elements exist, they are equal. However, an element of a semigroup (finite or infinite) may have several semigroup inverses, or no semigroup inverse at all.

1.3 Groups

A monoid is a *group* if each of its elements has a group inverse. A slightly weaker condition can be given.

Proposition 1.3. *A monoid is a group if and only if each of its elements has a right group inverse and a left group inverse.*

Proof. In a group, every element has a right group inverse and a left group inverse. Conversely, let G be a monoid in which every element has a right group inverse and a left group inverse. Let $g \in G$, let g' [g''] be a right [left] inverse of g. Thus, by definition, $gg' = 1$ and $g''g = 1$. It follows that $g'' = g''(gg') = (g''g)g' = g'$. Thus $g' = g''$ is an inverse of g. Thus G is a group. □

For finite monoids, this result can be further strengthened as follows:

Proposition 1.4. *A finite monoid G is a group if and only if every element of G has a left group inverse.*

Of course, the dual statement for right group inverses hold.

Proof. Let G be a finite monoid in which every element has a left group inverse. Given an element $g \in G$, consider the map $\varphi : G \to G$ defined by $\varphi(x) = gx$. We claim that φ is injective. Suppose that $gx = gy$ for some $x, y \in G$ and let g' be the left group inverse of g. Then $g'gx = g'gy$, that is $x = y$, proving the claim. Since G is finite, Proposition I.1.7 shows that φ is also surjective. In particular, there exists an element $g'' \in G$ such that $1 = gg''$. Thus every element of G has a right group inverse and by Proposition 1.3, G is a group. □

Proposition 1.5. *A group is a cancellative monoid. In a group, every element has a unique group inverse.*

Proof. Let G be a group. Let $g, x, y \in G$ and let g' be a group inverse of g. If $gx = gy$, then $g'gx = g'gy$, that is $x = y$. Similarly, $xg = yg$ implies $x = y$ and thus G is cancellative. In particular, if g' and g'' are two group inverses of g, $gg' = gg''$ and thus $g' = g''$. Thus every element has a unique inverse. □

In a group, the unique group inverse of an element x is denoted by x^{-1} and is called the *inverse* of x. Thus $xx^{-1} = x^{-1}x = 1$. It follows that the equation $gx = h$ $[xg = h]$ has a unique solution: $x = g^{-1}h$ $[x = hg^{-1}]$.

1.4 Ordered semigroups and monoids

An *ordered semigroup* is a semigroup S equipped with an order relation $\leqslant$ on S which is compatible with the product: for every $x, y \in S$, for every $u, v \in S^1$ $x \leqslant y$ implies $uxv \leqslant uyv$.

The notation $(S, \leqslant)$ will sometimes be used to emphasize the role of the order relation, but most of the time the order will be implicit and the notation S will be used for semigroups as well as for ordered semigroups. If $(S, \leqslant)$ is an ordered semigroup, then $(S, \geqslant)$ is also an ordered semigroup. *Ordered monoids* are defined analogously.

1.5 Semirings

A *semiring* is a quintuple consisting of a set k, two binary operations on k, written additively and multiplicatively, and two elements 0 and 1, satisfying the following conditions:

(1) k is a commutative monoid for the addition with identity 0,
(2) k is a monoid for the multiplication with identity 1,
(3) Multiplication is distributive over addition: for all $s, t_1, t_2 \in k$, $s(t_1 + t_2) = st_1 + st_2$ and $(t_1 + t_2)s = t_1s + t_2s$,
(4) for all $s \in k$, $0s = s0 = 0$.

A *ring* is a semiring in which the monoid $(k, +, 0)$ is a group. A semiring is *commutative* if its multiplication is commutative.

2 Examples

We give successively some examples of semigroups, monoids, groups, ordered monoids and semirings.

2.1 Examples of semigroups

(1) The set $\mathbb{N}_+$ of positive integers is a commutative semigroup for the usual addition of integers. It is also a commutative semigroup for the usual multiplication of integers.

(2) Let I and J be two nonempty sets. Define an operation on $I \times J$ by setting, for every $(i,j),(i',j') \in I \times J$,

$$(i,j)(i',j') = (i,j')$$

This defines a semigroup, usually denoted by $B(I,J)$.

(3) Let n be a positive integer. Let B_n be the set of all matrices of size $n \times n$ with zero-one entries and at most one nonzero entry. Equipped with the usual multiplication of matrices, B_n is a semigroup. For instance,

$$B_2 = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \right\}$$

This semigroup is nicknamed the *universal counterexample* because it provides many counterexamples in semigroup theory. Setting $a = \left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$ and $b = \left(\begin{smallmatrix} 0 & 0 \\ 1 & 0 \end{smallmatrix}\right)$, one gets $ab = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right)$, $ba = \left(\begin{smallmatrix} 0 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ and $0 = \left(\begin{smallmatrix} 0 & 0 \\ 0 & 0 \end{smallmatrix}\right)$. Thus $B_2 = \{a, b, ab, ba, 0\}$. Furthermore, the relations $aa = bb = 0$, $aba = a$ and $bab = b$ suffice to recover completely the multiplication in B_2.

(4) Let S be a set. Define an operation on S by setting $st = s$ for every $s,t \in S$. Then every element of S is a left zero, and S forms a *left zero semigroup*.

(5) Let S be the semigroup of matrices of the form

$$\begin{pmatrix} a & 0 \\ b & 1 \end{pmatrix}$$

where a and b are positive rational numbers, under matrix multiplication. We claim that S is a cancellative semigroup without identity. Indeed, since

$$\begin{pmatrix} a & 0 \\ b & 1 \end{pmatrix} \begin{pmatrix} x & 0 \\ y & 1 \end{pmatrix} = \begin{pmatrix} ax & 0 \\ bx+y & 1 \end{pmatrix}$$

it follows that if

$$\begin{pmatrix} a & 0 \\ b & 1 \end{pmatrix} \begin{pmatrix} x_1 & 0 \\ y_1 & 1 \end{pmatrix} = \begin{pmatrix} a & 0 \\ b & 1 \end{pmatrix} \begin{pmatrix} x_2 & 0 \\ y_2 & 1 \end{pmatrix}$$

then $ax_1 = ax_2$ and $bx_1 + y_1 = bx_2 + y_2$, whence $x_1 = x_2$ and $y_1 = y_2$, which proves that S is left cancellative. The proof that S is right cancellative is dual.

(6) If S is a semigroup, the set $\mathcal{P}(S)$ of subsets of S is also a semigroup, for the multiplication defined, for every $X, Y \in \mathcal{P}(S)$, by

$$XY = \{xy \mid x \in X, y \in Y\}$$

2.2 Examples of monoids

(1) The trivial monoid, denoted by 1, consists of a single element, the identity.
(2) The set $\mathbb{N}$ of nonnegative integers is a commutative monoid for the addition, whose identity is 0. It is also a commutative monoid for the max operation, whose identity is also 0 and for the multiplication, whose identity is 1.
(3) The monoid $U_1 = \{1, 0\}$ defined by its multiplication table $1 * 1 = 1$ and $0 * 1 = 0 * 0 = 1 * 0 = 0$.
(4) More generally, for each nonnegative integer n, the monoid U_n is defined on the set $\{1, a_1, \ldots, a_n\}$ by the operation $a_i a_j = a_j$ for each $i, j \in \{1, \ldots, n\}$ and $1a_i = a_i 1 = a_i$ for $1 \leqslant i \leqslant n$.
(5) The monoid $\tilde{U}_n$ has the same underlying set as U_n, but the multiplication is defined in the opposite way: $a_i a_j = a_i$ for each $i, j \in \{1, \ldots, n\}$ and $1a_i = a_i 1 = a_i$ for $1 \leqslant i \leqslant n$.
(6) The monoid B_2^1 is obtained from the semigroup B_2 by adding an identity. Thus $B_2^1 = \{1, a, b, ab, ba, 0\}$ where $aba = a$, $bab = b$ and $aa = bb = 0$.
(7) The *bicyclic monoid* is the monoid $M = \{(i, j) \mid (i, j) \in \mathbb{N}^2\}$ under the operation

$$(i, j)(i', j') = (i + i' - \min(j, i'), j + j' - \min(j, i'))$$

2.3 Examples of groups

(1) The set $\mathbb{Z}$ of integers is a commutative group for the addition, whose identity is 0.
(2) The set $\mathbb{Z}/n\mathbb{Z}$ of integers modulo n, under addition is also a commutative group.
(3) The set of 2×2 matrices with entries in $\mathbb{Z}$ and determinant ± 1 is a group under the usual multiplication of matrices. This group is denoted by $GL_2(\mathbb{Z})$.

2.4 Examples of ordered monoids

(1) Every monoid can be equipped with the equality order, which is compatible with the product. It is actually often convenient to consider a monoid M as the ordered monoid $(M, =)$.
(2) The natural order on nonnegative integers is compatible with addition and with the max operation. Thus $(\mathbb{N}, +, \leqslant)$ and $(\mathbb{N}, \max, \leqslant)$ are both ordered monoids.
(3) We let U_1^+ $[U_1^-]$ denote the monoid $U_1 = \{1, 0\}$ ordered by $1 < 0$ $[0 < 1]$.

2.5 Examples of semirings

(1) Rings are the first examples of semirings that come to mind. In particular, we let $\mathbb{Z}$, $\mathbb{Q}$ and $\mathbb{R}$, respectively, denote the rings of integers, rational and real numbers.
(2) The simplest example of a semiring which is not a ring is the *Boolean semiring* $\mathbb{B} = \{0, 1\}$ whose operations are defined by the following tables

$+$	0	1
0	0	1
1	1	1

$\times$	0	1
0	0	0
1	0	1

(3) If M is a monoid, then the set $\mathcal{P}(M)$ of subsets of M is a semiring with union as addition and multiplication given by

$$XY = \{xy \mid x \in X \text{ and } y \in Y\}$$

The empty set is the zero and the identity element is the singleton $\{1\}$.

(4) Other examples include the semiring of nonnegative integers $\mathbb{N} = (\mathbb{N}, +, \times)$ and its completion $\mathcal{N} = (\mathbb{N} \cup \{\infty\}, +, \times)$, where addition and multiplication are extended in the natural way

$$\begin{aligned} \text{for all } x \in \mathcal{N},\ & x + \infty = \infty + x = \infty \\ \text{for all } x \in \mathcal{N} - \{0\},\ & x \times \infty = \infty \times x = \infty \\ & \infty \times 0 = 0 \times \infty = 0 \end{aligned}$$

(5) The Min-Plus semiring is $\mathcal{M} = (\mathbb{N} \cup \{\infty\}, \min, +)$. This means that in this semiring the sum is defined as the minimum and the product as the usual addition. Note that ∞ is the zero of this semiring and 0 is its identity element.

3 Basic algebraic structures

3.1 Morphisms

On a general level, a morphism between two algebraic structures is a map preserving the operations. Therefore a *semigroup morphism* is a map φ from a semigroup S to a semigroup T such that, for every $s_1, s_2 \in S$,

(1) $\varphi(s_1 s_2) = \varphi(s_1)\varphi(s_2)$.

Similarly, a *monoid morphism* is a map φ from a monoid S to a monoid T satisfying (1) and

(2) $\varphi(1) = 1$.

A *morphism of ordered monoids* is a map φ from an ordered monoid $(S, \leqslant)$ to a monoid $(T, \leqslant)$ satisfying (1), (2) and, for every $s_1, s_2 \in S$ such that $s_1 \leqslant s_2$,

(3) $\varphi(s_1) \leqslant \varphi(s_2)$.

Formally, a *group morphism* between two groups G and G' is a monoid morphism φ satisfying, for every $s \in G$, $\varphi(s^{-1}) = \varphi(s)^{-1}$. In fact, this condition can be relaxed.

Proposition 3.6. *Let G and G' be groups. Then any semigroup morphism from G to G' is a group morphism.*

Proof. Let $\varphi : G \to G'$ be a semigroup morphism. Then by (1), $\varphi(1) = \varphi(1)\varphi(1)$ and thus $\varphi(1) = 1$ since 1 is the unique idempotent of G'. Thus φ is a monoid morphism. Furthermore, $\varphi(s^{-1})\varphi(s) = \varphi(s^{-1}s) = \varphi(1) = 1$ and similarly $\varphi(s)\varphi(s^{-1}) = \varphi(ss^{-1}) = \varphi(1) = 1$. □

A *semiring morphism* between two semirings k and k' is a map $\varphi : k \to k'$ which is a monoid morphism for the additive structure and for the multiplicative structure.

The semigroups [monoids, groups, ordered monoids, semirings], together with the morphisms defined above, form a category. We shall encounter in Chapter XVI another interesting category whose objects are semigroups and whose morphisms are called *relational morphisms*.

A morphism $\varphi : S \to T$ is an *isomorphism* if there exists a morphism $\psi : T \to S$ such that $\varphi \circ \psi = \mathrm{Id}_T$ and $\psi \circ \varphi = \mathrm{Id}_S$.

Proposition 3.7. *In the category of semigroups* [*monoids, groups, semirings*]*, a morphism is an isomorphism if and only if it is bijective.*

Proof. If $\varphi : S \to T$ an isomorphism, then φ is bijective since there exists a morphism $\psi : T \to S$ such that $\varphi \circ \psi = \mathrm{Id}_T$ and $\psi \circ \varphi = \mathrm{Id}_S$.

Suppose now that $\varphi : S \to T$ is a bijective morphism. Then φ^{-1} is a morphism from T to S, since, for each $x, y \in T$,

$$\varphi(\varphi^{-1}(x)\varphi^{-1}(y)) = \varphi(\varphi^{-1}(x))\varphi(\varphi^{-1}(y)) = xy$$

Thus φ is an isomorphism. □

Proposition 3.7 does not hold for morphisms of ordered monoids. In particular, if $(M, \leqslant)$ is an ordered monoid, the identity induces a bijective morphism from $(M, =)$ onto $(M, \leqslant)$ which is not in general an isomorphism. In fact, a morphism of ordered monoids $\varphi : M \to N$ is an isomorphism if and only if φ is a bijective monoid morphism and, for every $x, y \in M$, $x \leqslant y$ is equivalent with $\varphi(x) \leqslant \varphi(y)$.

Two semigroups [monoids, ordered monoids] are *isomorphic* if there exists an isomorphism from one to the other. As a general rule, we shall identify two isomorphic semigroups.

3.2 Subsemigroups

A *subsemigroup* of a semigroup S is a subset T of S such that $s_1 \in T$ and $s_2 \in T$ imply $s_1 s_2 \in T$. A *submonoid* of a monoid is a subsemigroup containing the identity. A *subgroup* of a group is a submonoid containing the inverse of each of its elements.

A subsemigroup G of a semigroup S is said to be a *group in* S if there is an idempotent $e \in G$ such that G, under the operation of S, is a group with identity e.

Proposition 3.8. *Let $\varphi : S \to T$ be a semigroup morphism. If S' is a subsemigroup of S, then $\varphi(S')$ is a subsemigroup of T. If T' is a subsemigroup of T, then $\varphi^{-1}(T')$ is a subsemigroup of S.*

Proof. Let $t_1, t_2 \in \varphi(S')$. Then $t_1 = \varphi(s_1)$ and $t_2 = \varphi(s_2)$ for some $s_1, s_2 \in S'$. Since S' is a subsemigroup of S, $s_1 s_2 \in S'$ and thus $\varphi(s_1 s_2) \in \varphi(S')$. Now since φ is a morphism, $\varphi(s_1 s_2) = \varphi(s_1)\varphi(s_2) = t_1 t_2$. Thus $t_1 t_2 \in \varphi(S')$ and $\varphi(S')$ is a subsemigroup of T.

Let $s_1, s_2 \in \varphi^{-1}(T')$. Then $\varphi(s_1), \varphi(s_2) \in T'$ and since T' is a subsemigroup of T, $\varphi(s_1)\varphi(s_2) \in T'$. Since φ is a morphism, $\varphi(s_1)\varphi(s_2) = \varphi(s_1 s_2)$ and thus $s_1 s_2 \in \varphi^{-1}(T')$. Therefore $\varphi^{-1}(T')$ is a subsemigroup of S. □

Proposition 3.8 can be summarised as follows: *substructures are preserved by morphisms and by inverses of morphisms.* A similar statement holds for monoid morphisms and for group morphisms.

3.3 Quotients and divisions

Let S and T be two semigroups [monoids, groups, ordered monoids]. Then T is a *quotient* of S if there exists a surjective morphism from S onto T.

Note that any ordered monoid $(M, \leqslant)$ is a quotient of the ordered monoid $(M, =)$, since the identity on M is a morphism of ordered monoid from $(M, =)$ onto $(M, \leqslant)$.

Finally, a semigroup T *divides* a semigroup S (notation $T \preccurlyeq S$) if T is a quotient of a subsemigroup of S.

Proposition 3.9. *The division relation is transitive.*

Proof. Suppose that $S_1 \preccurlyeq S_2 \preccurlyeq S_3$. Then there exists a subsemigroup T_1 of S_2, a subsemigroup T_2 of S_3 and surjective morphisms $\pi_1 : T_1 \to S_1$ and $\pi_2 : T_2 \to S_2$. Put $T = \pi_2^{-1}(T_1)$. Then T is a subsemigroup of S_3 and S_1 is a quotient of T since $\pi_1(\pi_2(T)) = \pi_1(T_1) = S_1$. Thus S_1 divides S_3. □

The next proposition shows that division is a partial order on finite semigroups, up to isomorphism.

Proposition 3.10. *Two finite semigroups that divide each other are isomorphic.*

Proof. We keep the notation of the proof of Proposition 3.9, with $S_3 = S_1$. Since T_1 is a subsemigroup of S_2 and T_2 is a subsemigroup of S_1, one has $|T_1| \leqslant |S_2|$ and $|T_2| \leqslant |S_1|$. Furthermore, since π_1 and π_2 are surjective, $|S_1| \leqslant |T_1|$ and $|S_2| \leqslant |T_2|$. It follows that $|S_1| = |T_1| = |S_2| = |T_2|$, whence $T_1 = S_2$ and $T_2 = S_1$. Furthermore, π_1 and π_2 are bijections and thus S_1 and S_2 are isomorphic. □

The term *division* stems from a property of finite groups, usually known as Lagrange's theorem. The proof is omitted but is not very difficult and can be found in any textbook on finite groups.

Proposition 3.11 (Lagrange)**.** *Let G and H be finite groups. If G divides H, then $|G|$ divides $|H|$.*

The use of the term *division* in semigroup theory is much more questionable since Lagrange's theorem does not extend to finite semigroups. However, this terminology is universally accepted and we shall stick to it.

Let us mention a few useful consequences of Lagrange's theorem.

Proposition 3.12. *Let G be a finite group. Then, for each $g \in G$, $g^{|G|} = 1$.*

Proof. Consider the set H of all powers of g. Since G is finite, two powers, say g^p and g^q with $q > p$, are equal. Since G is a group, it follows that $g^{q-p} = 1$. Let r be the smallest positive integer such that $g^r = 1$. Then $H = \{1, g, \ldots, g^{r-1}\}$ is a cyclic group and by Lagrange's theorem, $|G| = rs$ for some positive integer s. It follows that $g^{|G|} = (g^r)^s = 1^s = 1$. □

Proposition 3.13. *A nonempty subsemigroup of a finite group is a subgroup.*

Proof. Let G be a finite group and let S be a nonempty subsemigroup of G. Let $s \in S$. By Proposition 3.12, $s^{|G|} = 1$. Thus $1 \in S$. Consider now the map $\varphi : S \to S$ defined by $\varphi(x) = xs$. It is injective, for G is right cancellative, and hence bijective by Proposition I.1.7. Consequently, there exists an element s' such that $s's = 1$. Thus every element has a left inverse and by Proposition 1.4, S is a group. □

3.4 Products

Given a family $(S_i)_{i\in I}$ of semigroups [monoids, groups], the *product* $\Pi_{i\in I} S_i$ is the semigroup [monoid, group] defined on the cartesian product of the sets S_i by the operation

$$(s_i)_{i\in I}(s'_i)_{i\in I} = (s_i s'_i)_{i\in I}$$

Note that the semigroup 1 is the identity for the product of semigroups [monoids, groups]. Following a usual convention, which can also be justified in the framework of category theory, we put $\prod_{i\in\emptyset} S_i = 1$.

Given a family $(M_i)_{i\in I}$ of ordered monoids, the product $\prod_{i\in I} M_i$ is naturally equipped with the order

$$(s_i)_{i\in I} \leqslant (s'_i)_{i\in I} \text{ if and only if, for all } i \in I,\ s_i \leqslant s'_i.$$

The resulting ordered monoid is the product of the ordered monoids $(M_i)_{i\in I}$.

The next proposition, whose proof is obvious, shows that product preserves substructures, quotients and division. We state it for semigroups, but it can be readily extended to monoids and to ordered semigroups or monoids.

Proposition 3.14. *Let $(S_i)_{i\in I}$ and $(T_i)_{i\in I}$ be two families of semigroups such that, for each $i \in I$, S_i is a subsemigroup [quotient, divisor] of T_i. Then $\prod_{i\in I} S_i$ is a subsemigroup of [quotient, divisor] of $\prod_{i\in I} T_i$.*

3.5 Ideals

Let S be a semigroup. A *right ideal* of S is a subset R of S such that $RS \subseteq R$. Thus R is a right ideal if, for each $r \in R$ and $s \in S$, $rs \in R$. Symmetrically, a *left ideal* is a subset L of S such that $SL \subseteq L$. An *ideal* is a subset of S which is simultaneously a right and a left ideal.

Observe that a subset I of S is an ideal if and only if, for every $s \in I$ and $x, y \in S^1$, $xsy \in I$. Here, the use of S^1 instead of S allows us to include the cases $x = 1$ and $y = 1$, which are necessary to recover the conditions $SI \subseteq S$ and $IS \subseteq I$. Slight variations on the definition are therefore possible:

(1) R is a right ideal if and only if $RS^1 \subseteq R$ or, equivalently, $RS^1 = R$,

(2) L is a left ideal if and only if $S^1L \subseteq L$ or, equivalently, $S^1L = L$,

(3) I is an ideal if and only if $S^1IS^1 \subseteq I$ or, equivalently, $S^1IS^1 = I$.

Note that any intersection of ideals [right ideals, left ideals] of S is again an ideal [right ideal, left ideal].

Let R be a subset of a semigroup S. The ideal [right ideal, left ideal] *generated* by R is the set S^1RS^1 [RS^1, S^1R]. It is the smallest ideal [right ideal, left ideal] containing R. An ideal [right ideal, left ideal] is called *principal* if

it is generated by a single element. Note that the ideal [right ideal, left ideal] generated by an idempotent e is equal to SeS [eS, Se]. Indeed, the equality $S^1eS^1 = SeS$ follows from the fact that $e = eee$.

Ideals are stable under surjective morphisms and inverses of morphisms.

Proposition 3.15. *Let $\varphi : S \to T$ be a semigroup morphism. If J is an ideal of T, then $\varphi^{-1}(J)$ is a ideal of S. Furthermore, if φ is surjective and I is an ideal of S, then $\varphi(I)$ is an ideal of T. Similar results apply to right and left ideals.*

Proof. If J is an ideal of T, then

$$S^1\varphi^{-1}(J)S^1 \subseteq \varphi^{-1}(T^1)\varphi^{-1}(J)\varphi^{-1}(T^1) \subseteq \varphi^{-1}(T^1JT^1) \subseteq \varphi^{-1}(J)$$

Thus $\varphi^{-1}(J)$ is an ideal of S.

Suppose that φ is surjective. If I is an ideal of S, then

$$T^1\varphi(I)T^1 = \varphi(S^1)\varphi(I)\varphi(S^1) = \varphi(S^1IS^1) = \varphi(I)$$

Thus $\varphi(I)$ is an ideal of T. □

Let, for $1 \leqslant k \leqslant n$, I_k be an ideal of a semigroup S. The set

$$I_1I_2\cdots I_n = \{s_1s_2\cdots s_n \mid s_1 \in I_1, s_2 \in I_2, \ldots, s_n \in I_n\}$$

is the *product* of the ideals I_1, ..., I_n.

Proposition 3.16. *The product of the ideals $I_1, \ldots, I_n$ is an ideal contained in their intersection.*

Proof. Since I_1 and I_n are ideals, $S^1I_1 = I_1$ and $I_nS^1 = I_n$. Therefore

$$S^1(I_1I_2\cdots I_n)S^1 = (S^1I_1)I_2\cdots(I_nS^1) = I_1I_2\cdots I_n$$

and thus $I_1I_2\cdots I_n$ is an ideal. Furthermore, for $1 \leqslant k \leqslant n$, $I_1I_2\cdots I_n \subseteq S^1I_kS^1 = I_k$. Thus $I_1I_2\cdots I_n$ is contained in $\bigcap_{1\leqslant k\leqslant n} I_k$. □

A nonempty ideal I of a semigroup S is called *minimal* if, for every nonempty ideal J of S, $J \subseteq I$ implies $J = I$.

Proposition 3.17. *A semigroup has at most one minimal ideal.*

Proof. Let I_1 and I_2 be two minimal ideals of a semigroup S. Then by Proposition 3.16, I_1I_2 is a nonempty ideal of S contained in $I_1 \cap I_2$. Now since I_1 and I_2 are minimal ideals, $I_1I_2 = I_1 = I_2$. □

The existence of a minimal ideal is assured in two important cases, namely if S is finite or if S possesses a zero. In the latter case, 0 is the minimal ideal. A nonempty ideal $I \neq 0$ such that, for every nonempty ideal J of S, $J \subseteq I$ implies $J = 0$ or $J = I$ is called a 0-*minimal ideal*. It should be noted that a semigroup may have several 0-minimal ideals as shown in the next example.

Example 3.1. Let $S = \{s, t, 0\}$ be the semigroup defined by $xy = 0$ for every $x, y \in S$. Then 0 is the minimal ideal of S and $\{s, 0\}$ and $\{t, 0\}$ are two 0-minimal ideals.

3.6 Simple and 0-simple semigroups

A semigroup S is called *simple* if its only ideals are $\emptyset$ and S. It is called 0-*simple* if it has a zero, denoted by 0, if $S^2 \neq \{0\}$ and if $\emptyset$, 0 and S are its only ideals. The notions of *right simple*, *right* 0-*simple*, *left simple* and *left* 0-*simple* semigroups are defined analogously.

Lemma 3.18. *Let S be a 0-simple semigroup. Then $S^2 = S$.*

Proof. Since S^2 is a nonempty, nonzero ideal, one has $S^2 = S$. □

Proposition 3.19.

(1) *A semigroup S is simple if and only if $SsS = S$ for every $s \in S$.*

(2) *A semigroup S is 0-simple if and only if $S \neq \emptyset$ and $SsS = S$ for every $s \in S - 0$.*

Proof. We shall prove only (2), but the proof of (1) is similar.

Let S be a 0-simple semigroup. Then $S^2 = S$ by Lemma 3.18 and hence $S^3 = S$.

Let I be set of the elements s of S such that $SsS = 0$. This set is an ideal of S containing 0, but not equal to S since $\bigcup_{s\in S} SsS = S^3 = S$. Therefore $I = 0$. In particular, if $s \neq 0$, then $SsS \neq 0$, and since SsS is an ideal of S, it follows that $SsS = S$.

Conversely, if $S \neq \emptyset$ and $SsS = S$ for every $s \in S-0$, we have $S = SsS \subseteq S^2$ and therefore $S^2 \neq 0$. Moreover, if J is a nonzero ideal of S, it contains an element $s \neq 0$. We then have $S = SsS \subseteq SJS = J$, whence $S = J$. Therefore S is 0-simple. □

The structure of finite simple semigroups will be detailed in Section V.3.

3.7 Semigroup congruences

A *semigroup congruence* is a stable equivalence relation. Thus an equivalence relation $\sim$ on a semigroup S is a congruence if, for each $s, t \in S$ and $u, v \in S^1$, we have

$$s \sim t \text{ implies } usv \sim utv.$$

The set $S/\sim$ of equivalence classes of the elements of S is naturally equipped with the structure of a semigroup, and the function which maps every element onto its equivalence class is a semigroup morphism from S onto $S/\sim$. Four particular cases of congruences are extensively used.

(a) *Rees congruence*

Let I be an ideal of a semigroup S and let $\equiv_I$ be the equivalence relation identifying all the elements of I and separating the other elements. Formally, $s \equiv_I t$ if and only if $s = t$ or $s, t \in I$. Then $\equiv_I$ is a congruence called the *Rees congruence* of I. The quotient of S by $\equiv_I$ is usually denoted by S/I. The support of this semigroup is the set $(S - I) \cup 0$ and the multiplication (here denoted by $*$) is defined by

$$s * t = \begin{cases} st & \text{if } s, t, st \in S - I \\ 0 & \text{otherwise.} \end{cases}$$

(b) *Syntactic congruence*

Let P be a subset of a semigroup S. The *syntactic congruence* of P is the congruence $\sim_P$ over S defined by $s \sim_P t$ if and only if, for every $x, y \in S^1$,

$$xsy \in P \Longleftrightarrow xty \in P$$

The quotient semigroup $S/\sim_P$ is called the *syntactic semigroup* of P in S. The syntactic semigroup is particularly important in the theory of formal languages.

(c) *Congruence generated by a relation*

Let R be a relation on S, that is, a subset of $S \times S$. The set of all congruences containing R is nonempty since it contains the universal relation on S. Furthermore, it is closed under intersection. It follows that the intersection of all congruences containing R is a congruence, called the *congruence generated by* R. Let us give a more constructive definition.

Proposition 3.20. *The congruence generated by a symmetric relation R on a semigroup S is the reflexive-transitive closure of the relation*

$$\{(xry, xsy) \mid (r, s) \in R \text{ and } x, y \in S^1\}$$

Proof. If a congruence contains R, it certainly contains the relation

$$\bar{R} = \{(xry, xsy) \mid (r, s) \in R \text{ and } x, y \in S^1\}$$

and hence its reflexive-transitive closure $\bar{R}^*$. Therefore, it suffices to show that $\bar{R}^*$ is a congruence. Let $(u, v) \in \bar{R}^*$. By definition, there exists a finite sequence $u = u_0, u_1, \ldots, u_n = v$ such that

$$(u_0, u_1) \in \bar{R},\ (u_1, u_2) \in \bar{R},\ \cdots, (u_{n-1}, u_n) \in \bar{R}$$

Therefore, one has, for some $x_i, y_i, r_i, s_i \in S$ such that $(r_i, s_i) \in R$,

$$(u_0, u_1) = (x_0 r_0 y_0, x_0 s_0 y_0),\ (u_1, u_2) = (x_1 r_1 y_1, x_1 s_1 y_1),\ \cdots,$$
$$(u_{n-1}, u_n) = (x_{n-1} r_{n-1} y_{n-1}, x_{n-1} s_{n-1} y_{n-1})$$

Let now $x, y \in S^1$. Then the relations

$$(xu_i y, xu_{i+1} y) = (xx_i r_i y_i y, xx_i s_i y_i y) \in \bar{R} \qquad (0 \leqslant i \leqslant n-1)$$

show that $(xuy, xvy) \in \bar{R}^*$. Thus $\bar{R}^*$ is a congruence. □

(d) *Nuclear congruence*

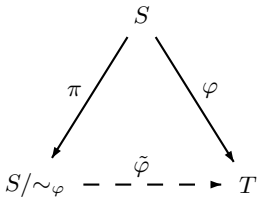
For each semigroup morphism $\varphi : S \to T$, the equivalence $\sim_\varphi$ defined on S by

$$x \sim_\varphi y \text{ if and only if } \varphi(x) = \varphi(y)$$

is a congruence. This congruence, called the *nuclear congruence* of φ, has the following standard property.

Proposition 3.21 (First isomorphism theorem)**.** *Let $\varphi : S \to T$ be a morphism of semigroups and let $\pi : S \to S/\sim_\varphi$ be the quotient morphism. Then there exists a unique semigroup morphism $\tilde{\varphi} : S/\sim_\varphi \to T$ such that $\varphi = \tilde{\varphi} \circ \pi$. Moreover, $\tilde{\varphi}$ is an isomorphism from $S/\sim_\varphi$ onto $\varphi(S)$.*

Proof. The situation is summed up in the following diagram:



Uniqueness is clear: if s is the $\sim_\varphi$-class of some element x, then necessarily

$$\tilde{\varphi}(s) = \varphi(x) \tag{3.1}$$

Furthermore, if x and y are arbitrary elements of s, then $\varphi(x) = \varphi(y)$. Therefore, there is a well-defined function $\tilde{\varphi}$ defined by Formula (3.1). Moreover, if $\pi(x_1) = s_1$ and $\pi(x_2) = s_2$, then $\pi(x_1x_2) = s_1s_2$, whence

$$\tilde{\varphi}(s_1)\tilde{\varphi}(s_2) = \varphi(x_1)\varphi(x_2) = \varphi(x_1x_2) = \tilde{\varphi}(s_1s_2)$$

Therefore $\tilde{\varphi}$ is a morphism. We claim that $\tilde{\varphi}$ is injective. Indeed, suppose that $\tilde{\varphi}(s_1) = \tilde{\varphi}(s_2)$, and let $x_1 \in \pi^{-1}(s_1)$ and $x_2 \in \pi^{-1}(s_2)$. Then $\varphi(x_1) = \varphi(x_2)$ and thus $x_1 \sim_\varphi x_2$. It follows that $\pi(x_1) = \pi(x_2)$, that is, $s_1 = s_2$. Thus $\tilde{\varphi}$ induces an isomorphism from $S/\sim_\varphi$ onto $\varphi(S)$. □

When two congruences are comparable, the quotient structures associated with them can also be compared.

Proposition 3.22 (Second isomorphism theorem)**.** *Let $\sim_1$ and $\sim_2$ be two congruences on a semigroup S and π_1 $[\pi_2]$ the canonical morphism from S onto $S/\sim_1$ $[S/\sim_2]$. If $\sim_2$ is coarser than $\sim_1$, there exists a unique surjective morphism $\pi : S/\sim_1 \to S/\sim_2$ such that $\pi \circ \pi_1 = \pi_2$.*

Proof. Since $\pi \circ \pi_1 = \pi_2$, Corollary I.1.13 shows that π is necessarily equal to the relation $\pi_2 \circ \pi_1^{-1}$. Furthermore, Proposition I.1.15 shows that this relation is actually a function.

Since π_1 and π_2 are morphisms,

$$\pi(\pi_1(s)\pi_1(t)) = \pi(\pi_1(st)) = \pi_2(st) = \pi_2(s)\pi_2(t) = \pi(\pi_1(s))\pi(\pi_1(t))$$

and thus π is a morphism. □

Proposition 3.23. *Let S be a semigroup, $(\sim_i)_{i\in I}$ be a family of congruences on S and $\sim$ be the intersection of these congruences. Then the semigroup $S/\sim$ is isomorphic to a subsemigroup of the product $\prod_{i\in I} S/\sim_i$.*

Proof. Denote by $\pi_i : S \to S/\sim_i$ the projections and by $\pi : S \to \prod_{i\in I} S/\sim_i$ the morphism defined by $\pi(s) = (\pi_i(s))_{i\in I}$ for every $s \in S$. The nuclear congruence of π is equal to $\sim$, and thus, by Proposition 3.21, $S/\sim$ is isomorphic to $\pi(S)$. □

Corollary 3.24. *Let S be a semigroup with zero having (at least) two distinct 0-minimal ideals I_1 and I_2. Then S is isomorphic to a subsemigroup of $S/I_1 \times S/I_2$.*

Proof. Under these assumptions, the intersection of I_1 and I_2 is 0 and thus the intersection of the Rees congruences $\equiv_{I_1}$ and $\equiv_{I_2}$ is the identity. It remains to apply Proposition 3.23 to conclude. □

(e) *Congruences of ordered semigroups*

Let $(M, \leqslant)$ be an ordered monoid. A *congruence of ordered semigroups* on M is a stable preorder coarser than $\leqslant$. If $\preceq$ is a congruence of ordered monoids, the equivalence $\sim$ associated with $\preceq$ is a semigroup congruence and the preorder $\preceq$ induces a stable order relation on the quotient monoid $S/\sim$, that will also be denoted by $\leqslant$. Finally, the function which maps each element onto its equivalence class is a morphism of ordered monoids from M onto $(M/\sim, \leqslant)$.

4 Transformation semigroups

4.1 Definitions

A [*partial*] *transformation* on a set P is a [partial] function from P to itself. A *permutation* is a bijective transformation. For instance, if $P = \{1, 2, 3\}$ and f, g and h are defined in the table below, then f is a transformation, g is a partial transformation and h is permutation.

	1	2	3
f	2	1	2
g	–	2	3
h	1	2	3

Let P be a set and S be a semigroup. A *right action* from S on P is a map $P \times S \to P$, denoted by $(p, s) \mapsto p \cdot s$, such that, for each $s, t \in S$ and $p \in P$,

$$(p \cdot s) \cdot t = p \cdot (st)$$

This condition implies that one may use the notation $p \cdot st$ in the place of $(p \cdot s) \cdot t$ or $p \cdot (st)$ without any ambiguity. We will follow this convention in the sequel.

An action is *faithful* if the condition

$$\text{for all } p \in P,\ p \cdot s = p \cdot t$$

implies $s = t$. A *transformation semigroup* on P is a semigroup S equipped with a faithful action of S on P.

Given an action of S on P, the relation $\sim$ defined on S by $s \sim t$ if and only if

$$\text{for all } p \in P,\ p \cdot s = p \cdot t$$

is a congruence on S and the action of S on P induces a faithful action of $S/\sim$ on P. The resulting transformation semigroup $(P, S/\sim)$ is called the transformation semigroup induced by the action of S on P.

Example 4.1. Each semigroup S defines a transformation semigroup (S^1, S), given by the faithful action $q \cdot s = qs$.

Example 4.2. A *right-zero semigroup* is a semigroup R in which the operation is defined by $st = t$ for every $s, t \in R$. One can associate to R a transformation semigroup (R, R) defined by the action $r \cdot s = s$ for every $r, s \in R$. In particular, if $R = \{1, \ldots, n\}$, this transformation semigroup is usually denoted by $\bar{\mathbf{n}}$.

A transformation semigroup (P, S) is said to be *fixed-point-free* if, for every $p \in P$ and every $s \in S$,

$$p \cdot s = p \quad \text{implies} \quad s = s^2.$$

For instance, translations of the plane form a fixed-point-free transformation semigroup.

4.2 Full transformation semigroups and symmetric groups

The *full transformation semigroup* on a set P is the semigroup $\mathcal{T}(P)$ of all transformations on P. If $P = \{1, \ldots, n\}$, the notation $\mathcal{T}_n$ is also used. According to the definition of a transformation semigroup, the product of two transformations α and β is the transformation $\alpha\beta$ defined by $p \cdot (\alpha\beta) = (p \cdot \alpha) \cdot \beta$. At this stage, the reader should be warned that the product $\alpha\beta$ is not equal to $\alpha \circ \beta$, but to $\beta \circ \alpha$. In other words, the operation on $\mathcal{T}(P)$ is reverse composition.

The set of all partial transformations on P is also a semigroup, denoted by $\mathcal{F}(P)$. Finally, the *symmetric group* on a set P is the group $\mathcal{S}(P)$ of all permutations on P. If $P = \{1, \ldots, n\}$, the notations $\mathcal{F}_n$ and $\mathcal{S}_n$ are also used.

The importance of these examples stems from the following embedding results.

Proposition 4.25. *Every semigroup S is isomorphic to a subsemigroup of the monoid $\mathcal{T}(S^1)$. In particular, every finite semigroup is isomorphic to a subsemigroup of $\mathcal{T}_n$ for some n.*

Proof. Let S be a semigroup. We associate with each element s of S the transformation on S^1, also denoted by s, and defined, for each $q \in S^1$, by $q \cdot s = qs$. This defines an injective morphism from S to $\mathcal{T}(S^1)$ and thus S is isomorphic to a subsemigroup of $\mathcal{T}(S^1)$. □

A similar proof leads to the following result, known as Cayley's theorem.

Theorem 4.26 (Cayley's theorem)**.** *Every group G is isomorphic to a subgroup of $\mathcal{S}(G)$. In particular, every finite group is isomorphic to a subgroup of $\mathcal{S}_n$ for some n.*

4.3 Product and division

Let $(P_i, S_i)_{i \in I}$ be a family of transformation semigroups. The *product* of this family is the transformation semigroup $(\prod_{i \in I} P_i, \prod_{i \in I} S_i)$. The action is defined componentwise:

$$(p_i)_{i \in I} \cdot (s_i)_{i \in I} = (p_i \cdot s_i)_{i \in I}.$$

A transformation semigroup (P, S) *divides* a transformation semigroup (Q, T) if there exists a surjective partial function $\pi \colon Q \to P$ and, for every $s \in S$, an element $\hat{s} \in T$, called a *cover* of s, such that, for each $q \in \mathrm{Dom}(\pi)$, $\pi(q) \cdot s = \pi(q \cdot \hat{s})$. The chosen terminology is justified by the following result.

Proposition 4.27. *If (P, S) divides (Q, T), then S divides T. If S divides T, then (S^1, S) divides (T^1, T).*

Proof. If (P, S) divides (Q, T), there exists a surjective partial function $\pi\colon Q \to P$ such that every element $s \in S$ has at least one cover. Furthermore, if $\hat{s}_1$ is a cover of s_1 and $\hat{s}_2$ is a cover of s_2, then $\hat{s}_1\hat{s}_2$ is a cover of s_1s_2, since, for each $q \in \mathrm{Dom}(\pi)$,

$$\pi(q)\cdot s_1s_2 = \pi(q\cdot \hat{s}_1)\cdot s_2 = \pi((q\cdot \hat{s}_1)\cdot \hat{s}_2) = \pi(q\cdot \hat{s}_1\hat{s}_2).$$

Therefore, the set of all covers of elements of S form a subsemigroup R of T. Furthermore, if two elements s_1 and s_2 have the same cover $\hat{s}$, then, for each $q \in \mathrm{Dom}(\pi)$,

$$\pi(q)\cdot s_1 = \pi(q\cdot \hat{s}) = \pi(q)\cdot s_2$$

Since π is surjective and the action of S is faithful, it follows that $s_1 = s_2$. Therefore, there is a well-defined map $\hat{s} \to s$ from R onto S and this map is a morphism.

Suppose now that S divides T. Then there exists a subsemigroup R of T and a surjective morphism π from R onto S, which can be extended to a surjective partial function from T^1 onto S^1, by setting $\pi(1) = 1$ if R is not a monoid. For each $s \in S$, choose an element $\hat{s} \in \pi^{-1}(s)$. Then, for every $q \in R^1$, $\pi(q\cdot \hat{s}) = \pi(q)s$ and thus (S^1, S) divides (T^1, T). □

5 Generators

5.1 A-generated semigroups

Given a subset A of a semigroup S, the subsemigroup of S *generated* by A is the smallest subsemigroup of S containing A. It is denoted by $\langle A\rangle$ and consists of all products $a_1 \cdots a_n$ of elements of A.

If S is a monoid, the *submonoid generated* by A is defined in a similar way, but it always contains the identity of S. Finally, if S is a group, the *subgroup generated* by A is the smallest subgroup of S containing A. It consists of all products of the form $a_1 \cdots a_n$, where each a_i is either an element of A or the inverse of an element of A.

A semigroup [monoid, group] generated by A is also called *A-generated*. For instance, the semigroup B_2 is A-generated, with $A = \{a, b\}$ and $a = \left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$ and $b = \left(\begin{smallmatrix} 0 & 0 \\ 1 & 0 \end{smallmatrix}\right)$.

A semigroup [monoid, group] is called *monogenic* if it is generated by a single element. More generally, a semigroup is said to be *finitely generated* if it admits a finite set of generators.

5.2 Cayley graphs

Given an A-generated monoid M, the *right Cayley graph* of M has M as set of vertices and as edges the triples of the form (m, a, ma), with $m \in M$ and $a \in A$. The *left Cayley graph* of M has the same set of vertices, but the edges are the triples of the form (m, a, am), with $m \in M$ and $a \in A$. Therefore the notion of Cayley graph depends on the given set of generators. However, it is a common

practice to speak of the Cayley graph of a monoid when the set of generators is understood.

The right and left Cayley graphs of B_2^1 are represented in Figure 5.1.

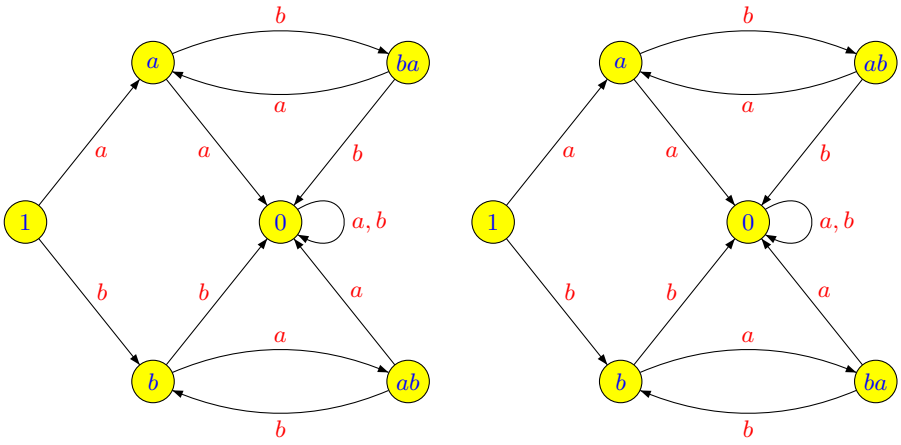


Figure 5.1. The left (on the left) and right (on the right) Cayley graph of B_2^1.

5.3 Free semigroups

Let A be a set called an *alphabet*, whose elements are called *letters*. A finite sequence of elements of A is called a *finite word* on A, or just a *word*. We denote the sequence $(a_0, a_1, \ldots, a_n)$ by mere juxtaposition

$$a_0 a_1 \cdots a_n.$$

The set of words is endowed with the operation of the *concatenation product* also called *product*, which associates with two words $x = a_0 a_1 \cdots a_p$ and $y = b_0 b_1 \cdots b_q$ the word $xy = a_0 a_1 \cdots a_p b_0 b_1 \cdots b_q$. This operation is associative. It has an identity, the *empty word*, denoted by 1 or by ε, which is the empty sequence.

We let A^* denote the set of words on A and by A^+ the set of nonempty words. The set A^* $[A^+]$, equipped with the concatenation product is thus a monoid with identity 1 [a semigroup]. The set A^* is called the *free monoid* on A and A^+ the *free semigroup* on A.

Let $u = a_0 a_1 \cdots a_p$ be a word in A^* and let a be a letter of A. A nonnegative integer i is said to be an *occurrence* of the letter a in u if $a_i = a$. We let $|u|_a$ denote the number of occurrences of a in u. Thus, if $A = \{a, b\}$ and $u = abaab$, one has $|u|_a = 3$ and $|u|_b = 2$. The sum

$$|u| = \sum_{a \in A} |u|_a$$

is the *length* of the word u. Thus $|abaab| = 5$.

5.4 Universal properties

Free structures are defined in category theory by a so-called universal property. The next proposition, which shows that A^+ $[A^*]$ satisfies this universal property,

justifies our terminology.

Proposition 5.28. *Let φ be a function from A to a semigroup [monoid] S. Then there exists a unique semigroup [monoid] morphism $\bar{\varphi} : A^+ \to S$ [$A^* \to S$] such that, for each $a \in A$, $\bar{\varphi}(a) = \varphi(a)$. Moreover, $\bar{\varphi}$ is surjective if and only if the set $\varphi(A)$ generates S.*

Proof. Define a mapping $\bar{\varphi} : A^+ \to S$ by setting, for each word $a_0a_1 \cdots a_n$,

$$\bar{\varphi}(a_0a_1 \cdots a_n) = \varphi(a_0)\varphi(a_1) \cdots \varphi(a_n)$$

One can easily verify that $\bar{\varphi}$ is the required morphism. On the other hand, any morphism $\bar{\varphi}$ such that $\bar{\varphi}(a) = \varphi(a)$ for each $a \in A$ must satisfy this equality, which shows it is unique.

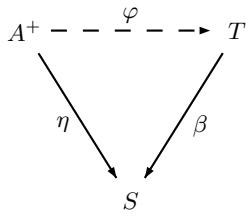
By construction, the set $\varphi(A)$ generates $\bar{\varphi}(A)$. Consequently, the morphism $\bar{\varphi}$ is surjective if and only if the set $\varphi(A)$ generates S. □

This result has several frequently used corollaries.

Corollary 5.29. *Let S be a semigroup and let A be a subset of S generating S. The identity map from A to S induces a morphism of semigroups from A^+ onto S.*

This morphism is called the *natural* morphism from A^+ onto S.

Corollary 5.30. *Let $\eta : A^+ \to S$ be a morphism and $\beta : T \to S$ be a surjective morphism. Then there exists a morphism $\varphi : A^+ \to T$ such that $\eta = \beta \circ \varphi$.*



Proof. Let us associate with each letter $a \in A$ an element $\varphi(a)$ of $\beta^{-1}(\eta(a))$. We thus define a function $\varphi : A \to T$, which, by Proposition 5.28, can be extended to a morphism $\varphi : A^+ \to T$ such that $\eta = \beta \circ \varphi$. □

5.5 Presentations and rewriting systems

A *semigroup presentation* is a pair $\langle A \mid R \rangle$, where A is an alphabet and R a subset of $A^+ \times A^+$. The elements of A are called *generators* and the ones of R, *relations*. The semigroup *presented by* $\langle A \mid R \rangle$ is the quotient of the free semigroup A^+ by the congruence $\sim_R$ generated by R. In other words, it is the semigroup generated by the set A subject to the relations R. This intuitive meaning is suggested by the notation. Indeed $\langle X \rangle$ traditionally denotes the semigroup generated by a set X and the vertical bar used as a separator can be interpreted as "such that", as in a definition like $\{n \in \mathbb{N} \mid \text{n is prime}\}$.

By extension, a semigroup is said to be *defined by a presentation* $\langle A \mid R \rangle$ if it is isomorphic to the semigroup presented by $\langle A \mid R \rangle$. Usually, we write $u = v$ instead of $(u, v) \in R$.

Monoid presentations are defined analogously by replacing the free semigroup by the free monoid. In particular, relations of the form $u = 1$ are allowed.

Example 5.1. The monoid presented by $\langle \{a, b\} \mid ab = ba \rangle$ is isomorphic to the additive monoid $\mathbb{N}^2$.

Example 5.2. The monoid presented by $\langle \{a, b\} \mid ab = 1 \rangle$ is the bicyclic monoid defined in Section 2.2.

6 Idempotents in finite semigroups

If S is a monogenic semigroup, generated by a single element x, the set S consists of the successive powers of x. If S is infinite, it is isomorphic to the additive semigroup of positive integers. If S is finite, there exist integers $i, p > 0$ such that

$$x^{i+p} = x^i.$$

The minimal i and p with this property are called respectively the *index* and the *period* of x. The semigroup S then has $i + p - 1$ elements and its multiplicative structure is represented in Figure 6.1.

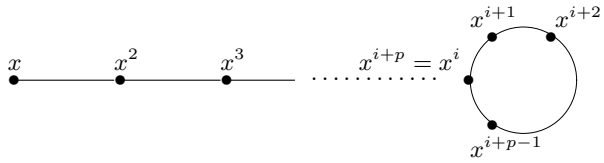


Figure 6.1. The semigroup generated by x.

The next result is a key property of finite semigroups.

Proposition 6.31. *Each element of a finite semigroup has an idempotent power.*

Proof. Let i and p be the index and the period of an element x. Observe that, for $k \geqslant i$, $x^k = x^{k+p}$. In particular, if k is a multiple of p, say $k = qp$, one has

$$(x^k)^2 = x^{2k} = x^{k+qp} = x^k$$

and thus x^k is idempotent. In fact, it is easy to see that the subsemigroup $\{x^i, \ldots, x^{i+p-1}\}$ is isomorphic to the additive group $\mathbb{Z}/p\mathbb{Z}$. □

Proposition 6.31 has two important consequences.

Corollary 6.32. *Every nonempty finite semigroup contains at least one idempotent.*

Proposition 6.33. *For each finite semigroup S, there exists an integer ω such that, for each $s \in S$, s^ω is idempotent.*

Proof. By Proposition 6.31, every element s of S has an idempotent power s^{n_s}. Let n be the least common multiple of the n_s, for $s \in S$. Then s^n is idempotent for each $s \in S$. □

The least integer ω satisfying the property stated in Proposition 6.33 is called the *exponent* of S.

Here is another elementary property connected with idempotents.

Proposition 6.34. *Let S be a finite semigroup and let $n = |S|$. For every sequence $s_1, \ldots, s_n$ of n elements of S, there exists an index $i \in \{1, \ldots, n\}$ and an idempotent $e \in S$ such that $s_1 \cdots s_i e = s_1 \cdots s_i$.*

Proof. Consider the sequence s_1, $s_1 s_2$, ..., $s_1 \cdots s_n$. If these elements are all distinct, the sequence exhausts the elements of S and one of them, say $s_1 \cdots s_i$, is idempotent. The result is thus clear in this case. Otherwise, two elements of the sequence are equal, say $s_1 \cdots s_i$ and $s_1 \cdots s_j$ with $i < j$. Then we have

$$s_1 \cdots s_i = s_1 \cdots s_i (s_{i+1} \cdots s_j) = s_1 \cdots s_i (s_{i+1} \cdots s_j)^\omega$$

where ω is the exponent of S. The proposition follows, since $(s_{i+1} \cdots s_j)^\omega$ is idempotent. □

If S is a semigroup and n is a positive integer, we set

$$S^n = \{s_1 \cdots s_n \mid s_i \in S \text{ for } 1 \leqslant i \leqslant n\}$$

Corollary 6.35. *Let S be a finite semigroup and let $E(S)$ be the set of idempotents of S. Then for every $n \geqslant |S|$, $S^n = SE(S)S$.*

We shall now present a more difficult result whose proof rests on a celebrated combinatorial theorem, due to Ramsey, which we shall admit without proof.

An *m-colouring* of a set E is a function from E to $\{1, \ldots, m\}$. An r-subset of E is a subset with r elements.

Theorem 6.36 (Ramsey)**.** *Let r, k, m be integers satisfying $k \geqslant r$, $m > 0$. Then there exists an integer $N = R(r, k, m)$ such that for every finite set E having at least N elements and for every m-colouring of the set of the r-subsets of E, there exists a k-subset of E of which all r-subsets have the same colour.*

The next result clearly generalises Proposition 6.34.

Theorem 6.37. *For each finite semigroup S, for each $k > 0$, there exists an integer $N > 0$ such that, for every alphabet A, for every morphism $\varphi : A^+ \to S$ and for every word w of A^+ of length greater than or equal to N, there exists an idempotent $e \in S$ and a factorisation $w = xu_1 \cdots u_k y$ with $x, y \in A^*$, $u_1, \ldots, u_k \in A^+$, $|xu_1 \cdots u_k| \leqslant N$ and $\varphi(u_1) = \ldots = \varphi(u_k) = e$.*

Proof. It suffices to prove the result for $k \geqslant 2$. Put $N = R(2, k+1, |S|)$ and let w be a word of length greater than or equal to N. Let $w = a_1 \cdots a_N w'$, where $a_1, \ldots, a_N$ are letters. We define a colouring of pairs of elements of $\{1, \ldots, N\}$ into $|S|$ colours in the following way: the colour of $\{i, j\}$, where $i < j$, is the element $\varphi(a_i \cdots a_{j-1})$ of S. According to Ramsey's theorem, one can find $k+1$ indices $i_0 < i_1 < \cdots < i_k$ such that all the pairs of elements of $\{i_0, \ldots, i_k\}$ have the same colour e. Since we assume $k \geqslant 2$, one has in particular

$$\begin{aligned} e = \varphi(a_{i_0}) \cdots \varphi(a_{i_1 - 1}) &= \varphi(a_{i_1}) \cdots \varphi(a_{i_2 - 1}) \\ &= \varphi(a_{i_0}) \cdots \varphi(a_{i_1 - 1}) \varphi(a_{i_1}) \cdots \varphi(a_{i_2 - 1}) = ee \end{aligned}$$

whereby $ee = e$. Thus e is idempotent and we obtain the required factorisation by taking $x = a_1 \cdots a_{i_0 - 1}$, $u_j = a_{i_{j-1}} \cdots a_{i_j - 1}$ for $1 \leqslant j \leqslant k$ and $y = a_{i_k} \cdots a_N w'$. □

There are many quantifiers in the statement of Theorem 6.37, but their order is important. In particular, the integer N does not depend on the size of A, which can even be infinite.

Proposition 6.38. *Let M be a finite monoid and let $\pi : A^* \to M$ be a surjective morphism. For any $n \geqslant 0$, there exists $N > 0$ and an idempotent e in M such that, for any $u_0, u_1, \ldots, u_N \in A^*$ there exists a sequence $0 \leqslant i_0 < i_1 < \ldots < i_n \leqslant N$ such that $\pi(u_{i_0} u_{i_0+1} \cdots u_{i_1-1}) = \pi(u_{i_1} u_{i_1+1} \cdots u_{i_2-1}) = \ldots = \pi(u_{i_{n-1}} \cdots u_{i_n-1}) = e$.*

Proof. Let $N = R(2, n+1, |M|)$ and let $u_0, u_1, \ldots, u_N$ be words of A^*. Let $E = \{1, \ldots, N\}$. We define a colouring into $|M|$ colours of the 2-subsets of E in the following way: the colour of the 2-subset $\{i, j\}$ (with $i < j$) is the element $\pi(u_i u_{i+1} \cdots u_{j-1})$ of M. According to Ramsey's theorem, one can find $k+1$ indices $i_0 < i_1 < \cdots < i_k$ such that all the 2-subsets of $\{i_0, \ldots, i_k\}$ have the same colour. In particular, since $k \geqslant 2$, one gets

$$\begin{aligned}\pi(u_{i_0} u_{i_0+1} \cdots u_{i_1-1}) = \pi(u_{i_1} u_{i_1+1} \cdots u_{i_2-1}) = \ldots = \pi(u_{i_{n-1}} \cdots u_{i_n-1}) \\ = \pi(u_{i_0} u_{i_0+1} \cdots u_{i_2-1})\end{aligned}$$

Let e be the common value of these elements. It follows from the equalities $\pi(u_{i_0} u_{i_0+1} \cdots u_{i_1-1}) = \pi(u_{i_1} u_{i_1+1} \cdots u_{i_2-1}) = \pi(u_{i_0} u_{i_0+1} \cdots u_{i_2-1})$ that $ee = e$ and thus e is idempotent. □

There is also a uniform version of Theorem 6.37, which is more difficult to establish.

Theorem 6.39. *For each finite semigroup S, for each $k > 0$, there exists an integer $N > 0$ such that, for every alphabet A, for every morphism $\varphi : A^+ \to S$ and for every word w of A^+ of length greater than or equal to N, there exists an idempotent $e \in S$ and a factorisation $w = x u_1 \cdots u_k y$ with $x, y \in A^*$, $|u_1| = \ldots = |u_k|$ and $\varphi(u_1) = \ldots = \varphi(u_k) = e$.*

Let us conclude this section with an important combinatorial result of Imre Simon. A *factorisation forest* is a function F that associates with every word x of A^2A^* a factorisation $F(x) = (x_1, \ldots, x_n)$ of x such that $n \geqslant 2$ and $x_1, \ldots, x_n \in A^+$. The integer n is the *degree* of the factorisation $F(x)$. Given a factorisation forest F, the *height function* of F is the function $h : A^* \to \mathbb{N}$ defined recursively by

$$h(x) = \begin{cases} 0 & \text{if } |x| \leqslant 1 \\ 1 + \max\{h(x_i) \mid 1 \leqslant i \leqslant n\} & \text{if } F(x) = (x_1, \ldots, x_n) \end{cases}$$

The *height* of F is the least upper bound of the heights of the words of A^*.

Let M be a finite monoid and let $\varphi : A^* \to M$ be a morphism. A factorisation forest F is *Ramseyan modulo* φ if, for every word x of A^2A^*, $F(x)$ is either of degree 2 or there exists an idempotent e of M such that $F(x) = (x_1, \ldots, x_n)$ and $\varphi(x_1) = \varphi(x_2) = \cdots = \varphi(x_n) = e$ for $1 \leqslant i \leqslant n$. The factorisation forest theorem was first proved by I. Simon in [147, 148, 149] and later improved in [29, 34, 35, 75]:

Theorem 6.40. *Let φ be a morphism from A^* to a finite monoid M. There exists a factorisation forest of height $\leqslant 3|M| - 1$ which is Ramseyan modulo φ.*

Proof. TO DO. □

7 Exercises

Section 2

Exercise 1. Show that, up to isomorphism, there are 5 semigroups of order 2. There are also 14 semigroups of order 3 and the number of semigroups of order $\leqslant 8$ is known to be 1843120128...

Section 3

Exercise 2. Let I and J be ideals of a semigroup S such that $I \subseteq J$. Show that I is an ideal of J, J/I is an ideal of S/I and $(S/I)/(J/I) = S/J$.

Exercise 3. Let T be a semigroup and let R and S be subsemigroups of T.

(1) Show that $R \cup S$ is a subsemigroup of T if and only if $RS \cup SR$ is a subset of $R \cup S$.
(2) Show that this condition is satisfied if R and S are both left ideals or both right ideals, or if either R or S is an ideal.
(3) Show that if R is an ideal of T, then $S \cap R$ is an ideal of S and $(S \cup R)/R = S/(S \cap R)$.

Exercise 4 (Hickey [59])**.** Let M be a monoid and let m be an element of M.

(1) Show that the operation $\circ$ defined on the set $mM \cap Mm$ by

$$xm \circ my = xmy$$

is well defined.
(2) Show that $(mM \cap Mm, \circ)$ is a monoid with identity m which divides M.

Section 4

Exercise 5. Show that, if $n \geqslant 2$, $\mathcal{T}_n$ is generated by the three functions a, b, and c of the table below

1	1	2	3	$\cdots$	$n-1$	n
a	2	3	4	$\cdots$	n	1
b	2	1	3	$\cdots$	$n-1$	n
c	1	2	3	$\cdots$	$n-1$	1

An element s of $\mathcal{T}_n$ is a function from $\{1, \ldots, n\}$ to itself. We define the *rank* $r(s)$ of s to be $|\operatorname{Im}(s)|$ and the defect of s to be $n - r(s)$. For instance, a and b have rank n and defect 0 and c has rank $n-1$ and defect 1.

Show that if s is an element of $\mathcal{T}_n$ of defect 1, then a, b and s generate $\mathcal{T}_n$.

Exercise 6. Let (P, S) be a transformation semigroup. Show that (P, S) divides $(P, 1_P) \times (S^1, S)$.

Section 5

Exercise 7. Show that the number of words of length n on a k-letter alphabet is k^n. Show that, if $k \geqslant 2$, the number of words of length $\leqslant n$ is $\frac{k^{n+1}-1}{k-1}$.

Exercise 8. Describe the semigroups $\langle \{a, b\} \mid a^2 = a, b^2 = b, ab = ba \rangle$ and $\langle \{a, b\} \mid a^2 = a, b^2 = b \rangle$.

Exercise 9. Show that the monoid presented by $\langle \{a, b\} \mid ab = ba = 1 \rangle$ is the group $(\mathbb{Z}, +)$.

Section 6

Exercise 10. Let S be a semigroup. Show that for each idempotent e of S, the sets

$$eS = \{es \mid s \in S\},\ Se = \{se \mid s \in S\} \text{ and } eSe = \{ese \mid s \in S\}$$

are subsemigroups of S. The semigroup eSe [eS, Se] is called the [*left, right*] local *semigroup* associated with e. Show that eSe is a monoid with identity e. Prove that an element s of S belongs to eSe [eS, Se] if and only if $es = s = se$ [$es = s$, $se = s$]. Show that $eSe = eS \cap Se$. What is the connection with Exercise 4?

Chapter III

Languages and automata

This chapter offers a brief overview of the theory of finite automata and formal languages.

There are different manners to describe a set of words, or a *language*. The *constructive approach* consists in giving a collection of basic languages and a set of construction rules to build new languages from previously defined ones. The definition of rational languages is of this type. In the *descriptive approach*, the words of a language are characterised by a property: the language of words of even length, the set of binary representations of prime numbers are examples of this approach. The *automata approach* is a special case of the descriptive approach: an automaton reads a word as input and decides whether the word is accepted or not. The set of words accepted by the automaton defines a language. Another variant of the descriptive approach consists in defining languages by logical formulas, a point of view further studied in Chapter IX.

The first three sections of this chapter review standard definitions on words, rational languages and finite automata. The minimal automaton of a language is defined in terms of morphisms of automata, which is more precise that just requiring that the number of states is minimal.

Section 4 gives an overview of standard constructions on automata related to various operations on languages: Boolean operations, product, star, quotients, inverses of morphisms.

The main result of Section 5 is Kleene's theorem, which states that a language is recognisable if and only if it is rational. Kleene's theorem is obtained in two steps. First, we present Glushkov's algorithm to pass from rational expressions to finite automata. Next, we use linear equations to pass from finite automata to rational expressions.

1 Words and languages

1.1 Words

Let $u = a_0a_1 \cdots a_n$ be a word of A^*. The *reversal* of u is the word $\tilde{u} = a_na_{n-1} \cdots a_0$ obtained by reading u from right to left.

A word $x \in A^*$ is a *factor* of u if $x = a_ra_{r+1} \cdots a_s$ for some r and s such that $0 \leqslant r \leqslant s \leqslant n$. This amounts to saying that there exist two words $v, w \in A^*$ such that $u = vxw$. Similarly, x is a *left factor*, or *prefix*, of u, if there exists

a word $w \in A^*$ such that $u = xw$ and x is a *right factor* or *suffix* of u, if there exists a word v of A^* such that $u = vx$.

Example 1.1. If $u = abaabba$, aba is a prefix, ba is a suffix and $abaa$, $baab$ are factors of u. One has $|u|_a = 4$ and $|u|_b = 3$, since u contains four occurrences of a and three of b.

1.2 Orders on words

The relation "being a prefix of" is an order relation on A^*, called the *prefix order*. which is partial if A contains at least two letters. There are other interesting orders on the free monoid. We describe two of them, the lexicographic order and the shortlex order.

Let A be an alphabet and let $\leqslant$ be a total order on A. The *lexicographic order* induced by $\leqslant$ is the total order on A^* used in a dictionary. Formally, it is the order $\leqslant_{\text{lex}}$ on A^* defined by $u \leqslant_{\text{lex}} v$ if and only if u is a prefix of v or $u = pau'$ and $v = pbv'$ for some $p \in A^*$, $a, b \in A$ with $a < b$. In the *shortlex order*, words are ordered by length and words of equal length are ordered according to the lexicographic order. Formally, it is the order $\leqslant$ on A^* defined by $u \leqslant v$ if and only if $|u| < |v|$ or $|u| = |v|$ and $u \leqslant_{\text{lex}} v$.

For instance, if $A = \{a, b\}$ with $a < b$, then $ababb <_{\text{lex}} abba$ but $abba < ababb$. The next proposition summarises elementary properties of the shortlex order. The proof is straightforward and omitted.

Proposition 1.1. *Let $u, v \in A^*$ and let $a, b \in A$.*

(1) If $u < v$, then $au < av$ and $ua < va$.

(2) If $ua \leqslant vb$, then $u \leqslant v$.

An important consequence of Proposition 1.1 is that the shortlex order is *stable*: if $u \leqslant v$, then $xuy \leqslant xvy$ for all $x, y \in A^*$.

1.3 Languages

Let A be a finite alphabet. Subsets of the free monoid A^* are called *languages*. For instance, if $A = \{a, b\}$, the sets $\{aba, babaa, bb\}$ and $\{a^n ba^n \mid n \geqslant 0\}$ are languages.

Several operations can be defined on languages. The *Boolean operations* comprise union, complement (with respect to the set A^* of all words), intersection and difference. Thus, if L and L' are languages of A^*, one has:

$$\begin{aligned} L \cup L' &= \{u \in A^* \mid u \in L \text{ or } u \in L'\} \\ L \cap L' &= \{u \in A^* \mid u \in L \text{ and } u \in L'\} \\ L^c &= A^* - L = \{u \in A^* \mid u \notin L\} \\ L - L' &= L \cap L'^c = \{u \in A^* \mid u \in L \text{ and } u \notin L'\} \end{aligned}$$

One can also mention the symmetric difference, defined as follows

$$L \triangle L' = (L - L') \cup (L' - L) = (L \cup L') - (L \cap L')$$

The *concatenation product* or simply *product* of two languages L and L' is the language

$$LL' = \{uu' \mid u \in L \text{ and } u' \in L'\}.$$

The product is an associative operation on the set of languages, which admits the language[1] $\{1\}$ as an identity, since the formula $\{1\}L = L\{1\} = L$ holds for each language L. The product is distributive over union, which means that, for all languages L, L_1 and L_2, one has

$$L(L_1 \cup L_2) = LL_1 \cup LL_2 \quad \text{and} \quad (L_1 \cup L_2)L = L_1L \cup L_2L \tag{1.1}$$

Therefore the languages over A^* form a semiring with union as addition and concatenation product as multiplication. For this reason, it is convenient to replace the symbol $\cup$ by $+$ and to let 0 denote the empty language and 1 the language $\{1\}$. For instance, (1.1) can be rewritten as $L(L_1 + L_2) = LL_1 + LL_2$ and $(L_1 + L_2)L = L_1L + L_2L$. Another convenient notation is to simply let u denote the language $\{u\}$. We shall use freely these conventions without any further warning.

Note that the product is not commutative if the alphabet contains at least two letters. Moreover, it is not distributive over intersection. For instance

$$(b \cap ba)\{a, aa\} = 0\{a, aa\} = 0 \quad \text{but}$$
$$b\{a, aa\} \cap ba\{a, aa\} = \{ba, baa\} \cap \{baa, baaa\} = baa$$

The powers of a language can be defined like in any monoid, by setting $L^0 = 1$, $L^1 = L$ and by induction, $L^n = L^{n-1}L$ for all $n > 0$. The *star* of a language L, denoted by L^*, is the sum (union) of all the powers of L:

$$L^* = \sum_{n \geqslant 0} L^n.$$

The operator L^+ is a variant of the star operator, obtained by considering the sum of all nonzero powers of a language:

$$L^+ = \sum_{n > 0} L^n.$$

Example 1.2. If $L = \{a, ba\}$ the words of L^+, ordered by increasing length, are a, aa, ba, aaa, aba, baa, $aaaa$, $aaba$, $abaa$, $baaa$, $baba$, $aaaaa$, $aaaba$, $aabaa$, $abaaa$, $ababa$, $baaaa$, $baaba$, $babaa$, etc.

Note that the notation A^* $[A^+]$ is compatible with the definition of the operations L^* and L^+. Also note the following formulas

$$0^* = 1, \quad 0^+ = 0 \quad \text{and} \quad 1^* = 1 = 1^+.$$

Let L be a language of A^* and let u be a word of A^*. The *left* [*right*] *quotient* $u^{-1}L$ $[Lu^{-1}]$ of L by u is defined as follows:

$$u^{-1}L = \{v \in A^* \mid uv \in L\} \quad \text{and} \quad Lu^{-1} = \{v \in A^* \mid vu \in L\}$$

It is easy to verify that the formula $v^{-1}(u^{-1}L) = (uv)^{-1}L$ and $u^{-1}(Lv^{-1}) = (u^{-1}L)v^{-1}$ hold for all words u and v of A^*. We also leave as an exercise to the reader the following formulas, where a is a letter:

$$u^{-1}(L_1 + L_2) = u^{-1}L_1 + u^{-1}L_2$$

[1] The language $\{1\}$, which consists of only one word, the empty word, should not be confused with the empty language.

$$u^{-1}(L_1 - L_2) = u^{-1}L_1 - u^{-1}L_2$$
$$u^{-1}(L_1 \cap L_2) = u^{-1}L_1 \cap u^{-1}L_2$$
$$a^{-1}(L_1L_2) = \begin{cases} (a^{-1}L_1)L_2 & \text{if } 1 \notin L_1, \\ (a^{-1}L_1)L_2 + a^{-1}L_2 & \text{if } 1 \in L_1 \end{cases}$$
$$a^{-1}L^* = (a^{-1}L)L^*$$

Example 1.3. Let $A = \{a, b\}$ and $L = A^*abaA^*$. Then

$$1^{-1}L = L \qquad a^{-1}L = A^*abaA^* + baA^*$$
$$b^{-1}L = L \qquad (ab)^{-1}L = A^*abaA^* + aA^*, \text{ etc.}$$

More generally, for any subset X of A^*, the left [right] quotient $X^{-1}L$ [LX^{-1}] of L by X is

$$X^{-1}L = \bigcup_{u \in X} u^{-1}L = \{v \in A^* \mid \text{ there exists } u \in X \text{ such that } uv \in L\}$$
$$LX^{-1} = \bigcup_{u \in X} Lu^{-1} = \{v \in A^* \mid \text{ there exists } u \in X \text{ such that } vu \in L\}$$

One has, for all languages X, Y and L (new exercises...)

$$(X + Y)^{-1}L = X^{-1}L + Y^{-1}L$$
$$(XY)^{-1}L = Y^{-1}(X^{-1}L)$$
$$(X^*)^{-1}L = L + (X^*)^{-1}(X^{-1}L)$$

To avoid too much parentheses, it is convenient to define precedence orders for operators on languages, summarised in Table 1.1.

Operator	**Priority**
L^*, L^c,	1
L_1L_2, $X^{-1}L$, LX^{-1}	2
$L_1 + L_2$, $L_1 \cap L_2$	3

Table 1.1. Operation precedence table.

The unary operators L^* and L^c have highest priority. The product and quotients have higher priority than union and intersection.

2 Rational languages

The set of *rational* (or *regular*) languages on A^* is the smallest set of languages $\mathcal{F}$ satisfying the following conditions:

(a) $\mathcal{F}$ contains the languages 0 and a for each letter $a \in A$,

(b) $\mathcal{F}$ is closed under finite union, product and star (i.e., if L and L' are languages of $\mathcal{F}$, then the languages $L + L'$, LL' and L^* are also in $\mathcal{F}$).

The set of rational languages of A^* is denoted by $\mathrm{Rat}(A^*)$.

This definition calls for a short comment. Indeed, there is a small subtlety in the definition, since one needs to ensure the existence of a "smallest set" satisfying the preceding conditions. For this, first observe that the set of all languages of A^* satisfies Conditions (a) and (b). Furthermore, the intersection of all the sets $\mathcal{F}$ satisfying Conditions (a) and (b) again satisfies these conditions: the resulting set is by construction the smallest set satisfying (a) and (b).

To obtain a more constructive definition, one can think of the rational languages as a kind of LEGO$^{\mathrm{TM}}$ box. The basic LEGO bricks are the empty language and the languages reduced to a single letter and three operators can be used to build more complex languages: finite union, product and star. For instance, it is easy to obtain a language consisting of a single word. If this word is the empty word, one makes use of the formula $0^* = 1$. For a word $a_1a_2\cdots a_n$ of positive length, one observes that

$$\{a_1a_2\cdots a_n\} = \{a_1\}\{a_2\}\cdots\{a_n\}.$$

Finite languages can be expressed as a finite union of singletons. For instance,

$$\{abaaba, ba, baa\} = abaaba + ba + baa$$

Consequently, finite languages are rational and the above definition is equivalent with the following more constructive version:

Proposition 2.2. *Let $\mathcal{F}_0$ be the set of finite languages of A^* and, for all $n > 0$, let $\mathcal{F}_{n+1}$ be the set of languages that can be written as $K + K'$, KK' or K^*, where K and K' are languages from $\mathcal{F}_n$. Then*

$$\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \mathcal{F}_2 \cdots$$

and the union of all the sets $\mathcal{F}_n$ is the set of rational languages.

Example 2.1. If $A = \{a, b\}$, the language $(a + ab + ba)^*$ is a rational language.

Example 2.2. The set L of all words containing a given factor u is rational, since $L = A^*uA^*$. Similarly, the set P of all words having the word p as a prefix is rational since $P = pA^*$.

Example 2.3. The set of words of even [odd] length is rational. Indeed, this language can be written as $(A^2)^*$ $[(A^2)^*A]$.

A variant of the previous description consists in using *rational expressions* to represent rational languages. Rational expressions are formal expressions (like polynomials in algebra or terms in logic) defined recursively as follows:

(1) 0 and 1 are rational expressions,

(2) for each letter $a \in A$, a is a rational expression,

(3) if e and e' are rational expressions, then $(e + e')$, (ee') and e^* are rational expressions.

In practice, unnecessary parentheses can be eliminated by applying the precedence rules given in Table 1.1. For instance, $((0^*a)(ba)^* + (bb^*))^*$ is a rational expression that should formally be written as $(((0^*a)(ba)^*) + (bb^*))^*$.

The *value* of a rational expression e, denoted by $v(e)$, is the language represented by e. The symbol 0 represents the empty language, the symbol 1 the

language reduced to the empty word, and each symbol a the language $\{a\}$. Finally, the operators union, product and star have their natural interpretation. Formally, one has

$$\begin{aligned} v(0) &= 0 \\ v(1) &= 1 \\ v(a) &= a \text{ for each letter } a \in A \\ v((e + e')) &= v(e) + v(e') \\ v((ee')) &= v(e)v(e') \\ v(e^*) &= (v(e))^* \end{aligned}$$

Beware not to confuse the notions of a rational expression and of a rational language. In particular, two rational expressions can represent the same language. For instance, the following expressions all represent the set of all words on the alphabet $\{a, b\}$.

$$e_1 = (a + b)^*, \quad e_2 = (a^*b)^*a^*, \quad e_3 = 1 + (a + b)(a + b)^*$$

The difficulty raised by this example is deeper than it seems. Even if a rational language can be represented by infinitely many different rational expressions, one could expect to have a unique reduced expression, up to a set of simple identities like $0 + L = L = L + 0$, $1L = L = L1$, $(L^*)^* = L^*$, $L + K = K + L$ or $L(K + K') = LK + LK'$. In fact, one can show there is no finite basis of identities for rational expressions: there exist no finite set of identities permitting to deduce all identities between rational expressions. Finding a complete infinite set of identities is already a difficult problem that leads to unexpected developments involving finite simple groups [36, 73].

We conclude this section by a standard result: rational languages are closed under morphisms. An extension of this result will be given in Proposition IV.1.1. The proof of this proposition can be found on page 76.

Proposition 2.3. *Let $\varphi : A^* \to B^*$ be a morphism. If L is a rational language of A^*, then $\varphi(L)$ is a rational language of B^*.*

3 Automata

3.1 Finite automata and recognisable languages

A finite *automaton* is a 5-tuple $\mathcal{A} = (Q, A, E, I, F)$, where Q is a finite set called the set of *states*, A is an alphabet, E is a subset of $Q \times A \times Q$, called the set of *transitions* and I and F are subsets of Q, called respectively the set of *initial states* and the set of *final states*.

It is convenient to represent an automaton by a labelled graph whose vertices are the states of the automaton and the edges represent the transitions. The initial [final] states are pictured by incoming [outgoing] arrows.

Example 3.1. Let $\mathcal{A} = (Q, A, E, I, F)$ where $Q = \{1, 2\}$, $I = \{1, 2\}$, $F = \{2\}$, $A = \{a, b\}$ and $E = \{(1, a, 1), (2, b, 1), (1, a, 2), (2, b, 2)\}$. This automaton is represented in Figure 3.1.

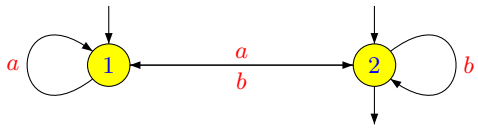


Figure 3.1. An automaton.

Two transitions (p, a, q) and (p', a', q') are *consecutive* if $q = p'$. A *path* in the automaton $\mathcal{A}$ is a finite sequence of consecutive transitions

$$c = (q_0, a_1, q_1), (q_1, a_2, q_2), \ldots, (q_{n-1}, a_n, q_n)$$

also denoted

$$c : q_0 \xrightarrow{a_1} q_1 \quad \cdots \quad q_{n-1} \xrightarrow{a_n} q_n \quad \text{or} \quad q_0 \xrightarrow{a_1 \cdots a_n} q_n.$$

The state q_0 is its *origin*, the state q_n its *end*, the word $a_1 \cdots a_n$ is its *label* and the integer n is its *length*. It is also convenient to consider that for each state $q \in Q$, there is an empty path $q \xrightarrow{1} q$ from q to q labelled by the empty word.

A path in $\mathcal{A}$ is called *initial* if its origin is an initial state and *final* if its end is a final state. It is *successful* (or *accepting*) if it is initial and final.

A state q is *accessible* if there is an initial path ending in q and it is *coaccessible* if there is a final path starting in q.

Example 3.2. Consider the automaton represented in Figure 3.1. The path

$$c : 1 \xrightarrow{a} 1 \xrightarrow{a} 2 \xrightarrow{b} 2 \xrightarrow{b} 1 \xrightarrow{a} 2 \xrightarrow{b} 2$$

is successful, since its end is a final state. However the path

$$c : 1 \xrightarrow{a} 1 \xrightarrow{a} 2 \xrightarrow{b} 2 \xrightarrow{b} 1 \xrightarrow{a} 2 \xrightarrow{b} 1$$

has the same label, but is not successful, since its end is 1, a nonfinal state.

A word is *accepted* by the automaton $\mathcal{A}$ if it is the label of at least one successful path (beware that it can be simultaneously the label of a nonsuccessful path). The *language recognised* (or *accepted*) by the automaton $\mathcal{A}$ is the set, denoted by $L(\mathcal{A})$, of all the words accepted by $\mathcal{A}$. Two automata are *equivalent* if they recognise the same language.

A language $L \subseteq A^*$ is *recognisable* if it is recognised by a finite automaton, that is, if there is a finite automaton $\mathcal{A}$ such that $L = L(\mathcal{A})$.

Example 3.3. Consider the automaton represented in Figure 3.2.

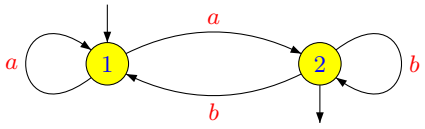


Figure 3.2. The automaton $\mathcal{A}$.

We let the reader verify that the language accepted by $\mathcal{A}$ is aA^*, the set of all words whose first letter is a.

Example 3.3 is elementary but it already raises some difficulties. In general, deciding whether a given word is accepted or not might be laborious, since a word might be the label of several paths. The notion of a deterministic automaton introduced in Section 3.2 permits one to avoid these problems.

We shall see in the Section 4 that the class of recognisable languages owns many convenient closure properties. We shall also see that the recognisable languages are exactly the rational languages. Before studying these results in more detail, it is good to realise that there are some nonrecognisable languages. One can establish this result by a simple, but nonconstructive argument (cf. Exercice 15). To get an explicit example, we shall establish a property of the recognisable languages known as the *pumping lemma*. Although this statement is formally true for any recognisable language, it is only interesting for the infinite ones.

Proposition 3.4 (Pumping lemma)**.** *Let L be a recognisable language. Then there is an integer $n > 0$ such that every word u of L of length greater than or equal to n can be factored as $u = xyz$ with $x, y, z \in A^*$, $|xy| \leqslant n$, $y \neq 1$ and, for all $k \geqslant 0$, $xy^k z \in L$.*

Proof. Let $\mathcal{A} = (Q, A, E, I, F)$ be an n-state automaton recognising L and let $u = a_1 \cdots a_r$ be a word of L of length $r \geqslant n$. Let $q_0 \xrightarrow{a_1} q_1 \cdots q_{r-1} \xrightarrow{a_r} q_r$ be a successful path labelled by u. As $r \geqslant n$, there are two integers i and j, with $i < j \leqslant n$, such that $q_i = q_j$. Therefore, the word $a_{i+1} \dots a_j$ is the label of a loop around q_i, represented in Figure 3.3.

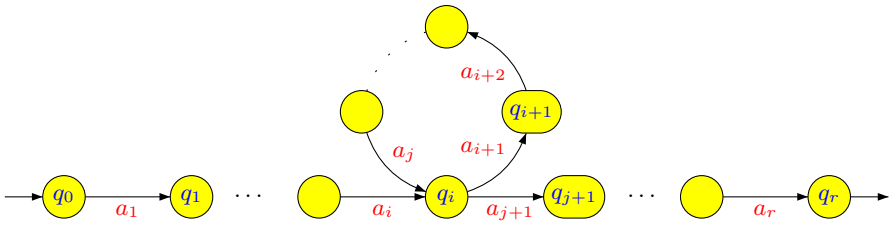


Figure 3.3. Illustration of the pumping lemma.

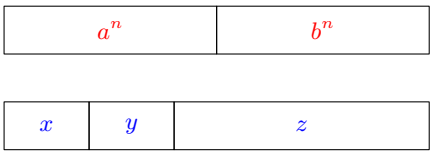
Let $x = a_1 \dots a_i$, $y = a_{i+1} \dots a_j$ and $z = a_{j+1} \dots a_r$. Then $|xy| \leqslant n$ and for all $k \geqslant 0$, one gets $xy^k z \in L$, since the word $xy^k z$ is the label of a successful path. □

Unfortunately, the pumping lemma does not characterise the recognisable languages. In other words, there exist some nonrecognisable languages which satisfy the pumping lemma (see Exercice 8). However, in some cases, the pumping lemma remains a convenient criterion to show that a language is not recognisable. Here is a standard example.

Corollary 3.5. *The language $\{a^n b^n \mid n \geqslant 0\}$ is not recognisable.*

Proof. Let $L = \{a^n b^n \mid n \geqslant 0\}$. Suppose that L is recognisable. By the pumping lemma, there is an integer $n > 0$ such that the word $a^n b^n$ can be written as $a^n b^n = xyz$ with $x, y, z \in A^*$, $|xy| \leqslant n$, $y \neq 1$ and $xy^2 z \in L$ (as one can see, we

only use a very weak form of the pumping lemma). As $|xy| \leqslant n$, the word xy is prefix of a^n and thus $x = a^r$, $y = a^s$ and $z = a^t b^n$, with $r+s+t = n$ and $s \neq 0$.



Then $xy^2z = a^r a^{2s} a^t b^n$ and since $s \neq 0$, one has $r + 2s + t \neq n$. It follows that xy^2z is not in L, a contradiction. Therefore L is not recognisable. □

3.2 Deterministic automata

An automaton $\mathcal{A} = (Q, A, E, I, F)$ is *deterministic* if I contains exactly one initial state and if, for every state $q \in Q$ and for every letter $a \in A$, there exists *at most* one state q' such that $q \xrightarrow{a} q'$ is a transition of E. When it exists, we let $q \cdot a$ denote this unique state q'. If q_- is the unique initial state, we adopt the notation $(Q, A, \cdot, q_-, F)$ instead of $(Q, A, E, \{q_-\}, F)$.

Example 3.4. The automaton represented in Figure 3.4 is deterministic.

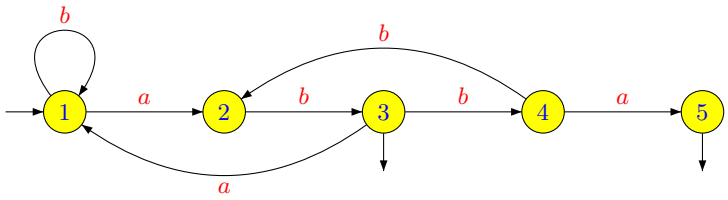


Figure 3.4. A deterministic automaton.

The following result is one of the cornerstones of automata theory. Its proof is based on a standard method known as the *powerset construction* or the *subset construction*[2].

Proposition 3.6. *Every finite automaton is equivalent to a deterministic one.*

Proof. Let $\mathcal{A} = (Q, A, E, I, F)$ be an automaton. Consider the deterministic automaton $D(\mathcal{A}) = (\mathcal{P}(Q), A, \cdot, I, \mathcal{F})$ where $\mathcal{F} = \{P \subseteq Q \mid P \cap F \neq \emptyset\}$ and, for each subset P of Q and for each letter $a \in A$,

$$P \cdot a = \{q \in Q \mid \text{there exists } p \in P \text{ such that } (p, a, q) \in E\}$$

We claim that $D(\mathcal{A})$ is equivalent to $\mathcal{A}$.

If $u = a_1 \cdots a_n$ is accepted by $\mathcal{A}$, there is a successful path

$$c : q_0 \xrightarrow{a_1} q_1 \quad \cdots \quad q_{n-1} \xrightarrow{a_n} q_n$$

The word u also defines a path

$$I = P_0 \xrightarrow{a_1} P_1 \quad \cdots \quad P_{n-1} \xrightarrow{a_n} P_n \tag{3.1}$$

[2]We shall give in Theorem IV.3.21 a purely algebraic proof of this result.

in $D(\mathcal{A})$. Let us show by induction on i that, for $0 \leqslant i \leqslant n$, $q_i \in P_i$. Since c is a successful path, one has $q_0 \in I = P_0$. Suppose that $q_{i-1} \in P_{i-1}$. Then since $q_{i-1} \xrightarrow{a_i} q_i$ is a transition, one gets $q_i \in P_{i-1} \cdot a_i = P_i$. For $i = n$, we get $q_n \in P_n$ and since c is a successful path, $q_n \in F$. It follows that P_n meets F and hence $P_n \in \mathcal{F}$. Therefore u is accepted by $D(\mathcal{A})$.

Conversely, let $u = a_1 \cdots a_n$ be a word accepted by $D(\mathcal{A})$ and let (3.1) be the successful path defined by u. Since P_n is a final state, one can choose an element q_n in $P_n \cap F$. We can now select, for $i = n, n-1, \ldots, 1$, an element q_{i-1} of P_{i-1} such that $q_{i-1} \xrightarrow{a_i} q_i$ is a transition in $\mathcal{A}$. Since $q_0 \in I$ and $q_n \in F$, the path $q_0 \xrightarrow{a_1} q_1 \cdots q_{n-1} \xrightarrow{a_n} q_n$ is successful, and thus u is accepted by $\mathcal{A}$. This proves the claim and the proposition. □

The subset construction converts a nondeterministic n-state automaton to a deterministic automaton with at most 2^n states. One can show that this bound is tight (see Exercise 10).

Example 3.5. Let $A = \{a, b\}$. Starting from the automaton $\mathcal{A}$ represented in Figure 3.5, we get the automaton $D(\mathcal{A})$ drawn in Figure 3.6. In practice, it suffices to compute the accessible states of $D(\mathcal{A})$, which gives the deterministic automaton shown in Figure 3.7.

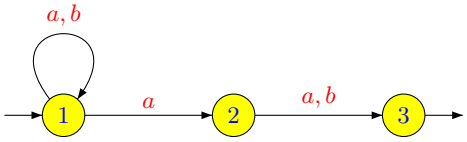


Figure 3.5. A nondeterministic automaton.

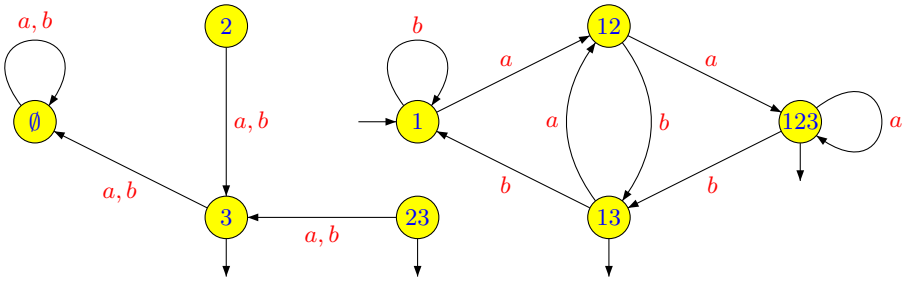


Figure 3.6. After determinisation...

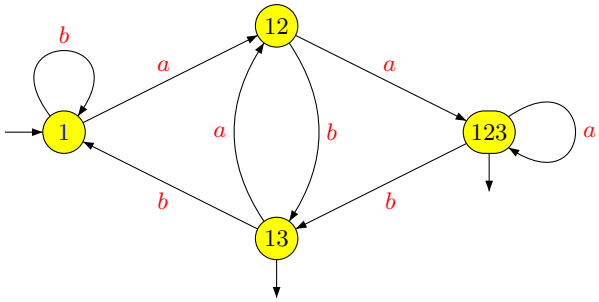


Figure 3.7. ... and trimming.

3.3 Complete, accessible, coaccessible and trimmed automata

An automaton $\mathcal{A} = (Q, A, \cdot, q_-, F)$ is *complete* if, for each state $q \in Q$ and for each letter $a \in A$, there is *at least* one state q' such that $q \xrightarrow{a} q'$ is a transition.

Example 3.6. The automaton represented in Figure 3.8 is neither complete, nor deterministic. It is not deterministic, since the transitions $(1, a, 1)$ and $(1, a, 2)$ have the same label and the same origin. It is not complete, since there is no transition of the form $2 \xrightarrow{a} q$.

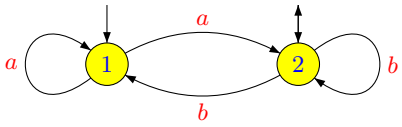


Figure 3.8. An incomplete, nondeterministic automaton.

On the other hand, the automaton represented in Figure 3.9 is complete and deterministic.

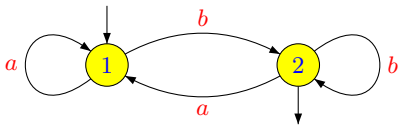


Figure 3.9. A complete and deterministic automaton.

An automaton is *accessible* if all its states are accessible. Similarly, it is *coaccessible* if all its states are coaccessible. Finally, an automaton is *trimmed* if it is simultaneously accessible and coaccessible. It is not difficult to see that every deterministic automaton is equivalent to a trimmed one.

3.4 Standard automata

The construction described in this section might look somewhat artificial, but it will be used in the study of the product and of the star operation.

A deterministic automaton is *standard* if there is no transition ending in the initial state.

Proposition 3.7. *Every deterministic automaton is equivalent to a deterministic standard automaton.*

Proof. Let $\mathcal{A} = (Q, A, E, q_-, F)$ be a deterministic automaton. If $\mathcal{A}$ is not standard, let p be a new state and $\mathcal{A}' = (Q \cup \{p\}, A, E', p, F')$ be the standard automaton defined by $E' = E \cup \{(p, a, q) \mid (q_-, a, q) \in E\}$ and

$$F' = \begin{cases} F & \text{if } q_- \notin F \\ F \cup \{p\} & \text{if } q_- \in F \end{cases}$$

Then the path $q_- \xrightarrow{a_0} q_1 \xrightarrow{a_1} q_2 \cdots q_{n-1} \xrightarrow{a_{n-1}} q_n$ is successful in $\mathcal{A}$ if and only if the path $p \xrightarrow{a_0} q_1 \xrightarrow{a_1} q_2 \cdots q_{n-1} \xrightarrow{a_{n-1}} q_n$ is successful in $\mathcal{A}'$. Consequently, $\mathcal{A}$ and $\mathcal{A}'$ are equivalent. □

Example 3.7. Standardisation is illustrated in Figure 3.10.

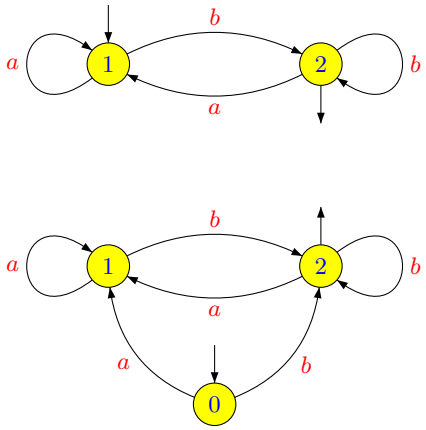


Figure 3.10. An automaton and its standardized version.

4 Operations on recognisable languages

We review in this section some classical results on finite automata. We give explicit constructions for the following operations: Boolean operations, product, star, quotients and inverses of morphisms.

4.1 Boolean operations

We give in this section the well known constructions for union, intersection and complement. Complementation is trivial, but requires a deterministic automaton.

Proposition 4.8. *The intersection of two recognisable languages is recognisable.*

Proof. Let L $[L']$ be a recognisable language of A^* recognised by the automaton $\mathcal{A} = (Q, A, E, I, F)$ $[\mathcal{A}' = (Q', A, E', I', F')]$. Consider the automaton

$$\mathcal{B} = (Q \times Q', A, T, I \times I', F \times F')$$

where

$$T = \Big\{ \big((q_1, q_1'), a, (q_2, q_2')\big) \mid (q_1, a, q_2) \in E \text{ and } (q_1', a, q_2') \in E' \Big\}.$$

A word $u = a_1 a_2 \cdots a_n$ is the label of a successful path in $\mathcal{B}$

$$(q_0, q_0') \xrightarrow{a_1} (q_1, q_1') \xrightarrow{a_2} (q_2, q_2') \quad \cdots \quad (q_{n-1}, q_{n-1}') \xrightarrow{a_n} (q_n, q_n')$$

if and only if the paths

$$q_0 \xrightarrow{a_1} q_1 \xrightarrow{a_2} q_2 \quad \cdots \quad q_{n-1} \xrightarrow{a_n} q_n$$

and

$$q_0' \xrightarrow{a_1} q_1' \xrightarrow{a_2} q_2' \quad \cdots \quad q_{n-1}' \xrightarrow{a_n} q_n$$

are successful paths of $\mathcal{A}$ and $\mathcal{A}'$ respectively. Therefore, $\mathcal{B}$ recognises $L \cap L'$. □

Example 4.1. If L $[L']$ is recognised by the automaton $\mathcal{A}$ $[\mathcal{A}']$ represented in Figure 4.1, then $L \cap L'$ is recognised by the automaton represented in Figure 4.2, or by its trimmed version (Figure 4.3).

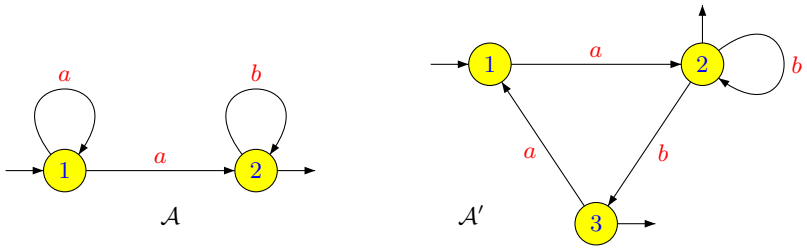


Figure 4.1. The automata $\mathcal{A}$ and $\mathcal{A}'$.

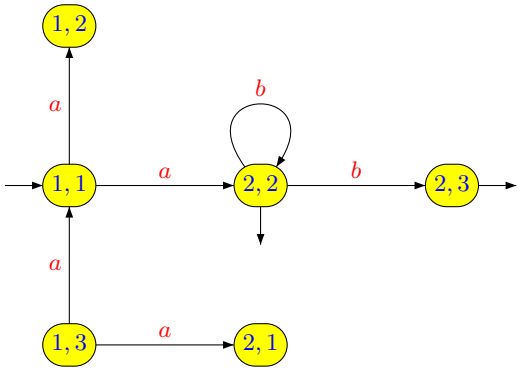


Figure 4.2. An automaton recognising $L \cap L'$.

In practice, one just computes the trimmed part of this automaton, represented in Figure 4.3.

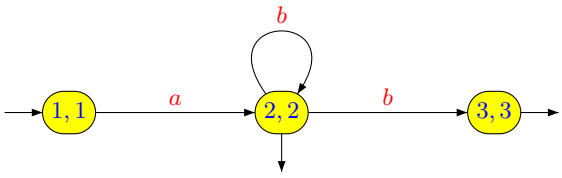


Figure 4.3. A trimmed automaton recognising $L \cap L'$.

The construction described in the proof of Proposition 4.8 has an advantage: if $\mathcal{A}$ and $\mathcal{A}'$ are deterministic automata, then the automaton $\mathcal{B}$ is also deterministic.

Proposition 4.9. *The union of two recognisable languages is recognisable.*

Proof. We give two different proofs of this result.

Proof using nondeterministic automata.

Let L $[L']$ be a recognisable language of A^* recognised by the automaton $\mathcal{A} = (Q, A, E, I, F)$ $[\mathcal{A}' = (Q', A, E', I', F')]$. Without loss of generality, one can suppose that Q and Q' are disjoint sets. Thus one can identify E and E' with subsets of $(Q \cup Q') \times A \times (Q \cup Q')$. Then $L + L'$ is recognised by the automaton $\mathcal{C} = (Q \cup Q', A, E \cup E', I \cup I', F \cup F')$.

Proof using deterministic automata.

Let L $[L']$ be a recognisable language of A^* recognised by the deterministic complete automaton $\mathcal{A} = (Q, A, \cdot, i, F)$ $[\mathcal{A}' = (Q', A, \cdot, i', F')]$. Consider the automaton

$$\mathcal{B} = (Q \times Q', A, \cdot, (i, i'), (F \times Q') \cup (Q \times F'))$$

where, for each letter $a \in A$, $(q, q') \cdot a = (q \cdot a, q' \cdot a)$. Let $u \in A^*$. Then $(i, i') \cdot u \in (F \times Q') \cup (Q \times F')$ if and only if $i \cdot u \in F$ or $i \cdot u \in F'$, that is, if and only if $u \in L$ or $u \in L'$. Thus $\mathcal{B}$ recognises $L \cup L'$. □

Example 4.2. If L $[L']$ is recognised by the automaton $\mathcal{A}$ $[\mathcal{A}']$ represented in Figure 4.1, then $L + L'$ is recognised by the automaton represented in Figure 4.4.

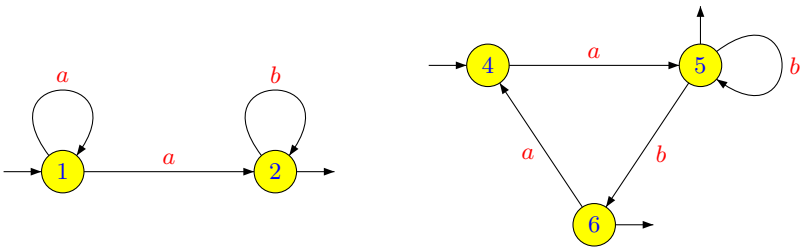


Figure 4.4. The automaton $\mathcal{C}$ recognising the union of L and L'.

Corollary 4.10. *Every finite language is recognisable.*

Proof. Since recognisable languages are closed under union, it suffices to verify that the singletons are recognisable. But it is clear that the language $a_1a_2\cdots a_n$ is recognised by the automaton represented in Figure 4.5. □

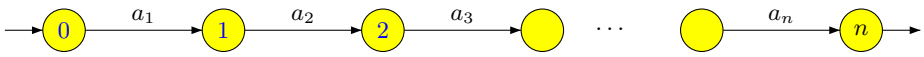


Figure 4.5. An automaton recognising $a_1 \cdots a_n$.

Proposition 4.11. *The complement of a recognisable language is recognisable.*

Proof. Let L be a recognisable language of A^* and let $\mathcal{A} = (Q, A, \cdot, q_-, F)$ be a complete deterministic automaton recognising L. Then the automaton $\mathcal{A}' = (Q, A, \cdot, q_-, Q - F)$ recognises L^c. Indeed, since $\mathcal{A}$ and $\mathcal{A}'$ are both deterministic and complete, every word u of A^* is the label of exactly one path starting in q_-. Let q be the end of this path. Then u belongs to L if and only if q belongs to F and u belongs to L^c if and only if q belongs to $Q - F$. □

Example 4.3. If L is recognised by the complete deterministic automaton $\mathcal{A}$, then L^c is recognised by the automaton $\mathcal{A}'$ represented in Figure 4.6.

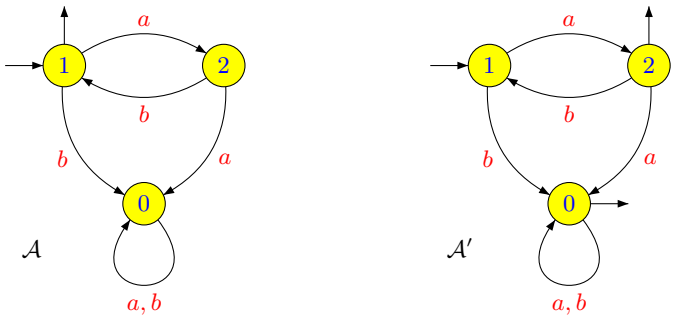


Figure 4.6. Complementation of a deterministic automaton.

4.2 Product

Proposition 4.12. *The product of two recognisable languages is recognisable.*

Proof. Let L_1 and L_2 be two recognisable languages of A^*, recognised by the automata $\mathcal{A}_1 = (Q_1, A, E_1, I_1, F_1)$ and $\mathcal{A}_2 = (Q_2, A, E_2, I_2, F_2)$, respectively. One may assume, by Propositions 3.6 and 3.7 that $\mathcal{A}_2$ is a standard deterministic automaton and thus in particular that $I_2 = \{i\}$. One can also suppose that Q_1 and Q_2 are disjoint. Let now $\mathcal{A} = (Q, A, E, I, F)$, where

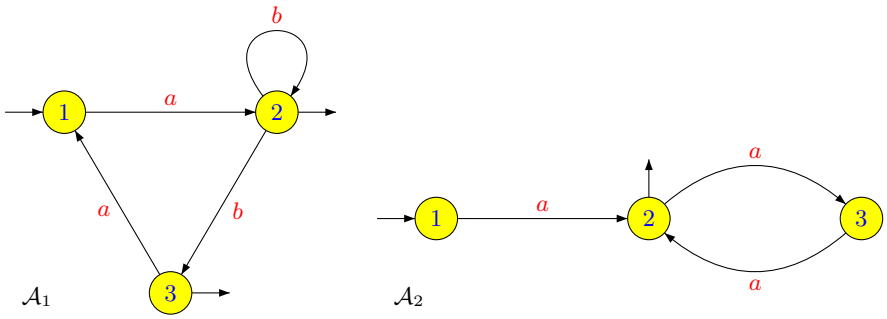
$$
\begin{aligned}
&Q = (Q_1 \cup Q_2) - \{i\},\\
&E = E_1 \cup \{(q, a, q') \in E_2 \mid q \neq i\} \cup \{(q_1, a, q') \mid q_1 \in F_1 \text{ and } (i, a, q') \in E_2\},\\
&I = I_1, \text{ and}
\end{aligned}
$$

$$F = \begin{cases} F_2 & \text{if } i \notin F_2, \\ F_1 \cup (F_2 - \{i\}) & \text{if } i \in F_2 \text{ (i.e. if } 1 \in L_2). \end{cases}$$

We claim that $\mathcal{A}$ recognises L_1L_2. If u is a word of L_1L_2, then $u = u_1u_2$ for some $u_1 \in L_1$ and $u_2 \in L_2$. Therefore, there is a successful path $c_1 : i_1 \xrightarrow{u_1} q_1$ in $\mathcal{A}_1$ (with $i_1 \in I_1$ and $q_1 \in F_1$) and a successful path $c_2 : i \xrightarrow{u_2} q_2$ in $\mathcal{A}_2$, with $q_2 \in F_2$. If $u_2 = 1$, then L_2 contains the empty word, the path c_1 is a successful path of $\mathcal{A}$ and u is accepted by $\mathcal{A}$. If u_2 is not the empty word, let a be the first letter of u_2 and let $i \xrightarrow{a} q$ be the first transition of c_2. Since $q_1 \in F_1$, $q_1 \xrightarrow{a} q$ is by definition a transition of E. Furthermore, if $q' \xrightarrow{b} q''$ is a transition of c_2 different from the first transition, then q' is the end of a transition of $\mathcal{A}_2$. Since $\mathcal{A}_2$ is standard, this implies $q' \neq i$ and it follows from the definition of E that the transition $q' \xrightarrow{b} q''$ is also a transition of $\mathcal{A}$. Let c_2' be the path of $\mathcal{A}$ obtained by replacing in c_2 the first transition $i \xrightarrow{a} q$ by $q_1 \xrightarrow{a} q$. The resulting path c_1c_2' is a successful path of $\mathcal{A}$ with label u and hence u is accepted by $\mathcal{A}$.

Conversely, let u be a word accepted by $\mathcal{A}$. Then u is the label of a successful path $c : i_1 \xrightarrow{u} f$ of $\mathcal{A}$. Since the initial states of $\mathcal{A}$ are contained in Q_1, and since there is no transition of $\mathcal{A}$ starting in Q_2 and ending in Q_1, c visits first some states of Q_1 and then possibly some states of Q_2. If all the states visited by c are in Q_1, one has in particular $f \in Q_1$. But this is only possible if $1 \in L_2$, and in this case, c is also a successful path of $\mathcal{A}_1$, and hence $u \in L_1 \subseteq L_1L_2$. If c visits some states of Q_2, then c contains a unique transition of the form $e = (q_1, a, q_2)$ with $q_1 \in F_1$ and $q_2 \in Q_2$. Therefore $c = c_1ec_2$, where c_1 is a path in $\mathcal{A}_1$ and c_2 is a path in $\mathcal{A}_2$. Denoting by u_1 [u_2] the label of c_1 [c_2], we get $u = u_1au_2$. Since c_1 is a successful path in $\mathcal{A}_1$, one has $u_1 \in L_1$. Moreover, by definition of E, $e' = (i, a, q_2)$ is a transition of $\mathcal{A}_2$. Therefore the path $e'c_2$ is a successful path in $\mathcal{A}_2$ with label au_2. It follows that $au_2 \in L_2$ and thus $u \in L_1L_2$, proving the claim and the proposition. □

Example 4.4. If L_1 [L_2] is recognised by the automaton $\mathcal{A}_1$ [$\mathcal{A}_2$], then L_1L_2 is recognised by the automaton $\mathcal{A}$ represented in Figure 4.7.



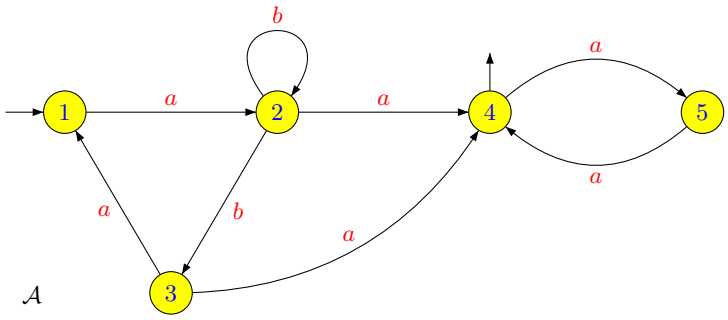


Figure 4.7. An automaton recognising L_1L_2.

4.3 Star

Proposition 4.13. *The star of a recognisable language is recognisable.*

Proof. Let L be a recognisable language of A^*, recognised by the deterministic standard automaton $\mathcal{A} = (Q, A, E, q_-, F)$. Let $\mathcal{A}' = (Q, A, E', \{q_-\}, F \cup \{q_-\})$ be the nondeterministic automaton defined by

$$E' = E \cup \{(q, a, q') \mid q \in F \text{ and } (q_-, a, q') \in E\}.$$

Let us show that $\mathcal{A}'$ recognises L^*. If u is a word of L^*, then either u is the empty word, which is accepted by $\mathcal{A}'$ since q_- is a final state, or $u = u_1u_2\cdots u_n$ with $u_1, \ldots, u_n \in L - 1$. Each u_i is the label of a successful path of $\mathcal{A}$, say $c_i : q_- \xrightarrow{u_i} q_i$ with $q_i \in F$. Let a_i be the first letter of u_i and let $q_- \xrightarrow{a_i} p_i$ be the first transition of c_i. Let $i \in \{2, \ldots, n\}$. As $q_{i-1} \in F$, the definition of E' shows that $q_{i-1} \xrightarrow{a_i} p_i$ is a transition of $\mathcal{A}'$. Denote by c'_i the path obtained by replacing in c_i the first transition $q_- \xrightarrow{a_i} p_i$ by $q_{i-1} \xrightarrow{a_i} p_i$. This defines, for $2 \leqslant i \leqslant n$, a path $c'_i : q_{i-1} \xrightarrow{u_i} q_i$ in $\mathcal{A}'$. Therefore, the path $c_1c'_2\cdots c'_n$ is a successful path of label u in $\mathcal{A}'$ and hence u is accepted by $\mathcal{A}'$.

Conversely, let u be a word accepted by $\mathcal{A}'$. If $u = 1$, one has $u \in L^*$. Otherwise, u is the label of a nonempty successful path c of $\mathcal{A}'$. This path can be factored as

$$c = q_- \xrightarrow{u_0} q_1 \xrightarrow{a_1} q'_1 \xrightarrow{u_1} q_2 \xrightarrow{a_2} q'_2 \quad \cdots \quad q_n \xrightarrow{a_n} q'_n \xrightarrow{u_n} q_{n+1}$$

where the transitions $e_1 = q_1 \xrightarrow{a_1} q'_1$, $e_2 = q_2 \xrightarrow{a_2} q'_2$, $\ldots$, $e_n = q_n \xrightarrow{u_n} q'_{n+1}$ are exactly the transitions of $E' - E$ occurring in c. Thus by definition of E', one gets, for $1 \leqslant i \leqslant n$, $q_i \in F$ and $e'_i = (q_-, a_i, q'_i) \in E$. Furthermore, $q_{n+1} \in F \cup \{q_-\}$ since c is a successful path. Consequently, the paths

$$q_- \xrightarrow{a_i} q'_i \xrightarrow{u_i} q_{i+1}$$

are paths of $\mathcal{A}$. For $1 \leqslant i \leqslant n-1$, these paths are successful, since $q_i \in F$. Moreover, since $\mathcal{A}$ is standard[3], q_{n+1} is different from q_- and hence $q_{n+1} \in F$. Consequently $a_iu_i \in L$ for $1 \leqslant i \leqslant n$. Since $q_- \xrightarrow{u_0} q_1$ is also a successful path of $\mathcal{A}$, one also has $u_0 \in L$, and hence $u \in L^*$. □

[3]This is the only place where this property is used

Example 4.5. If L is recognised by the standard deterministic automaton represented in Figure 4.8, then L^* is recognised by the nondeterministic automaton represented in Figure 4.9.

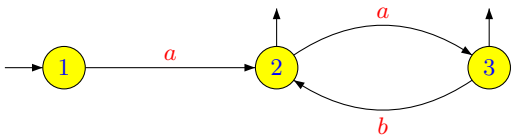


Figure 4.8. A standard automaton recognising L.

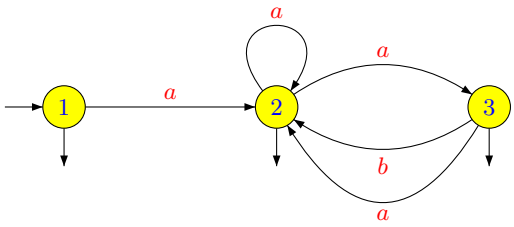


Figure 4.9. An automaton recognising L^*.

Example 4.6. This example shows that the algorithm above does not work if one does not start with a standard automaton. Indeed, consider the automaton $\mathcal{A}$ represented in Figure 4.10, which recognises the language $L = (ab)^*a$. Then the nondeterministic automaton $\mathcal{A}' = (Q, A, E', \{q_-\}, F \cup \{q_-\})$, where

$$E' = E \cup \{(q, a, q') \mid q \in F \text{ and } (q_-, a, q') \in E\}.$$

does not recognise L^*. Indeed, the word ab is accepted by $\mathcal{A}'$ but is not a word of L^*.

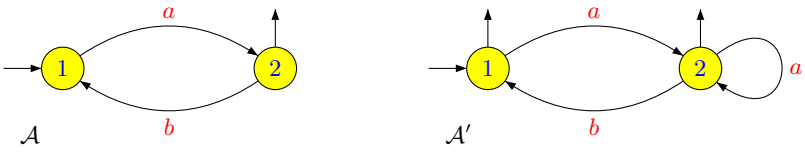


Figure 4.10. A nonstandard automaton $\mathcal{A}$ and the automaton $\mathcal{A}'$.

4.4 Quotients

We first treat the left quotient by a word and then the general case.

Proposition 4.14. *Let $\mathcal{A} = (Q, A, \cdot, q_-, F)$ be a deterministic automaton recognising a language L of A^*. Then, for each word u of A^*, the language $u^{-1}L$ is recognised by the automaton $\mathcal{A}_u = (Q, A, \cdot, q_- \cdot u, F)$, obtained from $\mathcal{A}$ by changing the initial state. In particular $u^{-1}L$ is recognisable.*

Proof. First the following formulas hold:

$$\begin{aligned} u^{-1}L &= \{v \in A^* \mid uv \in L\} \\ &= \{v \in A^* \mid q_- \cdot (uv) \in F\} \\ &= \{v \in A^* \mid (q_- \cdot u) \cdot v \in F\}. \end{aligned}$$

Therefore $u^{-1}L$ is accepted by $\mathcal{A}_u$. □

Proposition 4.15. *Any quotient of a recognisable language is recognisable.*

Proof. Let $\mathcal{A} = (Q, A, E, I, F)$ be an automaton recognising a language L of A^* and let K be a language of A^*. We do not assume that K is recognisable. Setting

$$I' = \{q \in Q \mid q \text{ is the end of an initial path whose label belongs to } K\}$$

we claim that the automaton $\mathcal{B} = (Q, A, E, I', F)$ recognises $K^{-1}L$. Indeed, if $u \in K^{-1}L$, there exists a word $x \in K$ such that $xu \in L$. Therefore, there is a successful path with label xu, say $p \overset{x}{\rightsquigarrow} q \overset{u}{\rightsquigarrow} r$. By construction, p is an initial state, r is a final state and q belongs to I'. It follows that u is accepted by $\mathcal{B}$.

Conversely, if a word u is accepted by $\mathcal{B}$, it is the label of a final path starting in a state q of I'. By definition of I', there is a word $x \in K$ which is the label of an initial path ending in q. Consequently, the word xu is accepted by $\mathcal{A}$ and hence belongs to L. It follows that $u \in K^{-1}L$, which proves the claim.

For the language LK^{-1}, a similar proof works by considering the automaton (Q, A, E, I, F'), where

$$F' = \{q \in Q \mid q \text{ is the origin of a final path whose label belongs to } K.\}$$ □

4.5 Inverses of morphisms

We now show that recognisable languages are closed under inverses of morphisms. A concise proof of an extension of this result will be given in Proposition IV.2.11.

Proposition 4.16. *Recognisable languages are closed under inverses of morphisms.*

Proof. Let $\varphi : A^* \to B^*$ be a morphism and let L be a recognisable language of B^*. Let $\mathcal{B} = (Q, B, E, I, F)$ be an automaton recognising L. Let $\mathcal{A} = (Q, A, T, I, F)$, with

$$T = \{(p, a, q) \mid \text{there exists a path of label } \varphi(a) \text{ from } p \text{ to } q \text{ in } \mathcal{B}\}$$

Let us show that the automaton $\mathcal{A}$ recognises $\varphi^{-1}(L)$. First, if u is accepted by $\mathcal{A}$, there exists a successful path in $\mathcal{A}$ labelled by u. Consequently, there exists a successful path in $\mathcal{B}$ labelled by $\varphi(u)$. Thus $\varphi(u)$ is accepted by $\mathcal{B}$ and $u \in \varphi^{-1}(L)$.

Let now $u = a_1 \cdots a_n$ be a word of $\varphi^{-1}(L)$. Since the word $\varphi(u)$ is accepted by L, there exists a successful path in $\mathcal{B}$ labelled by $\varphi(u)$. This path is obtained by concatenating paths of successive labels $\varphi(a_1), \ldots \varphi(a_n)$, say

$$q_0 \xrightarrow{\varphi(a_1)} q_1 \quad \cdots \quad q_{n-1} \xrightarrow{\varphi(a_n)} q_n$$

These paths in turn define a successful path in $\mathcal{A}$ labelled by u:

$$q_0 \xrightarrow{a_1} q_1 \quad \cdots \quad q_{n-1} \xrightarrow{a_n} q_n$$

which shows that u is accepted by $\mathcal{A}$. □

Example 4.7. Let $\varphi : \{a, b\}^* \to \{a, b, c\}^*$ be the morphism defined by $\varphi(a) = ac$ and $\varphi(b) = cba$. Let L be the language of $\{a, b, c\}^*$ recognised by the nondeterministic automaton $\mathcal{A}$ represented in Figure 4.11.

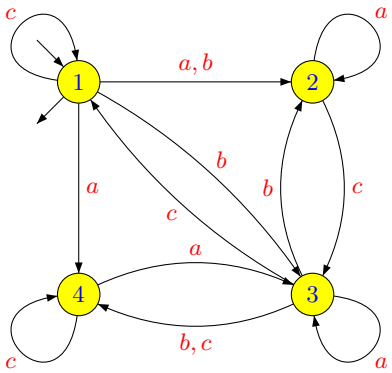


Figure 4.11. A nondeterministic automaton.

We first identify the paths labelled by ac and cba, as shown on the left-hand side of Figure 4.12 and then simply replace ac by a and cba by b to obtain the nondeterministic automaton on the right-hand side of Figure 4.12.

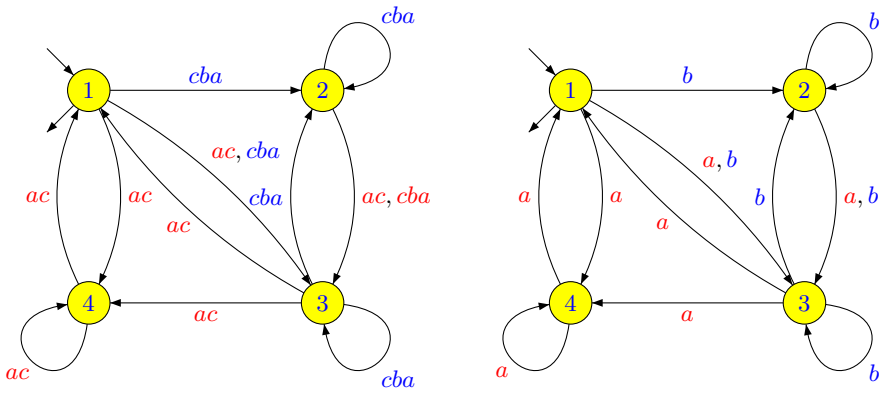


Figure 4.12. Computation of a nondeterministic automaton recognising $\varphi^{-1}(L)$.

4.6 Minimal automata

Let L be a language of A^*. The *Nerode automaton* of L is the deterministic automaton $\mathcal{A}(L) = (Q, A, \cdot, L, F)$ where $Q = \{u^{-1}L \mid u \in A^*\}$, $F = \{u^{-1}L \mid$

$u \in L\}$ and the transition function is defined, for each $a \in A$, by the formula

$$(u^{-1}L)\cdot a = a^{-1}(u^{-1}L) = (ua)^{-1}L$$

Beware of this rather abstract definition. Each state of $\mathcal{A}(L)$ is a left quotient of L by a word, and hence is a language of A^*. The initial state is the language L, and the set of final states is the set of all left quotients of L by a word of L.

Proposition 4.17. *A language L is recognisable if and only if the set $\{u^{-1}L \mid u \in A^*\}$ is finite. In this case, L is recognised by its Nerode automaton.*

Proof. Let L be a recognisable language, accepted by the deterministic automaton $\mathcal{A} = (Q, A, \cdot, q_-, F)$. By Proposition 4.14, the language $u^{-1}L$ is accepted by the automaton $\mathcal{A}_u = (Q, A, \cdot, q_- \cdot u, F)$. If n is the number of states of $\mathcal{A}$, there are at most n automata of the form $\mathcal{A}_u$ and hence at most n distinct languages of the form $u^{-1}L$.

Conversely, if the set $\{u^{-1}L \mid u \in A^*\}$ is finite, the Nerode automaton of L is finite and recognises L. Indeed, a word u is accepted by $\mathcal{A}(L)$ if and only if $L\cdot u = u^{-1}L$ is a final state, that is if $u \in L$. It follows that L is recognisable. □

Let us define a partial order on deterministic automata as follows. Let $\mathcal{A} = (Q, A, E, q_-, F)$ and $\mathcal{A}' = (Q', A, E', q'_-, F')$ be two deterministic automata. Then $\mathcal{A}' \leqslant \mathcal{A}$ if there is a surjective function $\varphi : Q \to Q'$ such that $\varphi(q_-) = q'_-$, $\varphi^{-1}(F') = F$ and, for every $u \in A^*$ and $q \in Q$, $\varphi(q\cdot u) = \varphi(q)\cdot u$.

Let L be a recognisable language. The next proposition shows that, amongst the accessible deterministic automata recognising L, the Nerode automaton of L is minimal for this partial order. For this reason it is called the *minimal accessible automaton* of L.

Proposition 4.18. *Let $\mathcal{A} = (Q, A, \cdot, q_-, F)$ be an accessible and complete deterministic automaton accepting L. For each state q of Q, let L_q be the language recognised by $(Q, A, \cdot, q, F)$. Then*

$$\mathcal{A}(L) = (\{L_q \mid q \in Q\}, A, \cdot, L_{q_-}, \{L_q \mid q \in F\}),$$

where, for all $a \in A$ and for all $q \in Q$, $L_q\cdot a = L_{q\cdot a}$. Furthermore, the map $q \mapsto L_q$ defines a morphism from $\mathcal{A}$ onto $\mathcal{A}(L)$.

Proof. Let q be a state of Q. Since q is accessible, there is a word u of A^* such that $q_- \cdot u = q$, and by Proposition 4.14, one has $L_q = u^{-1}L$. Conversely, if u is a word, one has $u^{-1}L = L_q$ with $q = q_- \cdot u$. Therefore

$$\{L_q \mid q \in Q\} = \{u^{-1}L \mid u \in A^*\} \quad \text{and} \quad \{L_q \mid q \in F\} = \{u^{-1}L \mid u \in L\}$$

which proves the first part of the statement.

We claim that the map $\varphi : q \mapsto L_q$ defines a morphism from $\mathcal{A}$ onto $\mathcal{A}(L)$. Indeed, for all $a \in A$, one has

$$\varphi(q\cdot a) = L_{q\cdot a} = L_q\cdot a = \varphi(q)\cdot a \tag{4.1}$$

which proves the claim. □

The direct computation of the Nerode automaton is probably the most efficient method for a computation by hand, because it directly gives the minimal automaton. In practice, one starts with the quotient $L = 1^{-1}L$ and one maintains a table of quotients of L. For each quotient R, it suffices to compute the quotients $a^{-1}R$ for each letter a. These quotients are compared to the existing list of quotients and possibly added to this list. But there is a hidden difficulty: the comparison of two rational expressions is not always easy since a given language might be represented by two very different rational expressions.

Example 4.8. For $L = (a(ab)^*)^* \cup (ba)^*$, we get successively:

$$
\begin{aligned}
1^{-1}L &= L = L_1\\
a^{-1}L_1 &= (ab)^*(a(ab)^*)^* = L_2\\
b^{-1}L_1 &= a(ba)^* = L_3\\
a^{-1}L_2 &= b(ab)^*(a(ab)^*)^* \cup (ab)^*(a(ab)^*)^* = bL_2 \cup L_2 = L_4\\
b^{-1}L_2 &= \emptyset\\
a^{-1}L_3 &= (ba)^* = L_5\\
b^{-1}L_3 &= \emptyset\\
a^{-1}L_4 &= a^{-1}(bL_2 \cup L_2) = a^{-1}L_2 = L_4\\
b^{-1}L_4 &= b^{-1}(bL_2 \cup L_2) = L_2 \cup b^{-1}L_2 = L_2\\
a^{-1}L_5 &= \emptyset\\
b^{-1}L_5 &= a(ba)^* = L_3
\end{aligned}
$$

which gives the minimal automaton represented in Figure 4.13.

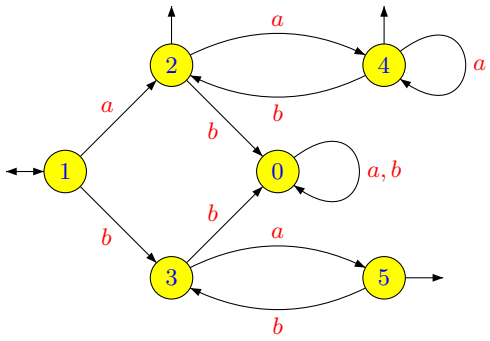


Figure 4.13. The minimal automaton of L.

There are standard algorithms for minimising a given trimmed deterministic automaton [64] based on the computation of the *Nerode equivalence*. Let $\mathcal{A} = (Q, A, E, q_-, F)$ be a trimmed deterministic automaton. The Nerode equivalence $\sim$ on Q is defined by $p \sim q$ if and only if, for every word $u \in A^*$,

$$p \cdot u \in F \iff q \cdot u \in F$$

One can show that $\sim$ is actually a congruence, which means two things: first, F is *saturated* by $\sim$, that is, $p \in F$ and $p \sim q$ implies $q \in F$, and secondly $p \sim q$

implies $p\cdot x \sim q\cdot x$ for all $x \in A^*$. It follows that there is a well-defined quotient automaton $\mathcal{A}/\!\sim = (Q/\!\sim, A, E, \tilde{q}_-, F/\!\sim)$, where $\tilde{q}_-$ is the equivalence class of q_-.

Proposition 4.19. *Let $\mathcal{A}$ be a trimmed [and complete] deterministic automaton. Then $\mathcal{A}/\!\sim$ is the minimal [complete] automaton of $\mathcal{A}$.*

We shall in particular use the following consequence.

Corollary 4.20. *A trimmed deterministic automaton is minimal if and only if its Nerode equivalence is the identity.*

One can show that the operations of completion and of minimisation commute. In other words, if $\mathcal{A}$ is a trimmed automaton and if $\bar{\mathcal{A}}$ is its completion, then the minimal automaton of $\bar{\mathcal{A}}$ is the completion of the minimal automaton of $\mathcal{A}$.

Example 4.9. The minimal and minimal complete automata of $(ab)^*$ are given in Figure 4.14.

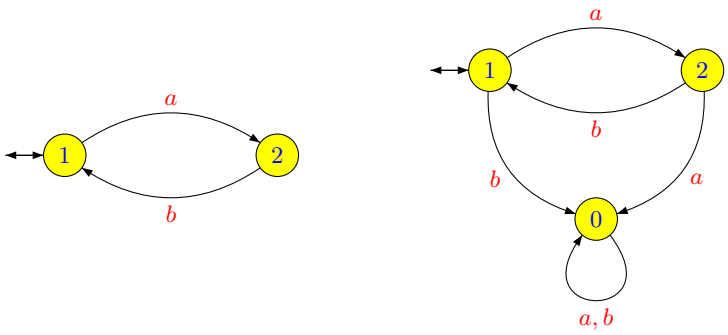


Figure 4.14. The minimal and minimal complete automata of $(ab)^*$.

Example 4.10. The minimal and minimal complete automata of aA^*b are given in Figure 4.15.

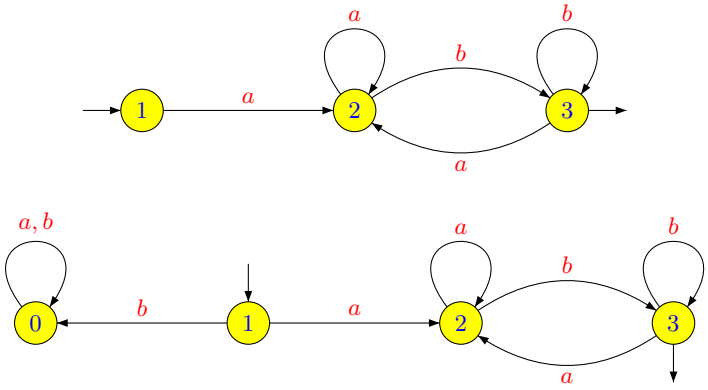


Figure 4.15. The minimal and minimal complete automata of aA^*b.

5 Rational versus recognisable

5.1 Local languages

A language L of A^* is said to be *local* if there exist two subsets P and S of A and a subset N of A^2 such that [4]

$$L - 1 = (PA^* \cap A^*S) - A^*NA^*.$$

For instance, if $A = \{a, b, c\}$, the language

$$(abc)^* = 1 \cup [(aA^* \cap A^*c) - A^*\{aa, ac, ba, bb, cb, cc)A^*]$$

is local. The terminology can be explained as follows: in order to check whether a nonempty word belongs to L, it suffices to verify that its first letter is in P, its last letter is in S and its factors of length 2 are not in N: all these conditions are local. Conversely, if a language L is local, it is easy to recover the parameters P, S and N. Indeed, P [S] is the set of first [last] letters of the words of L, and N is the set of words of length 2 that are factors of no word of L.

It is easy to compute a deterministic automaton recognising a local language, given the parameters P, S and N.

Proposition 5.21. *Let $L = (PA^* \cap A^*S) - A^*NA^*$ be a local language. Then L is recognised by the automaton $\mathcal{A}$ in which the set of states is $A \cup \{1\}$, the initial state is 1, the set of final states is S, and the transitions are given by the rules $1 \cdot a = a$ if $a \in P$ and $a \cdot b = b$ if $ab \notin N$.*

Proof. Let $u = a_1 \cdots a_n$ be a word accepted by $\mathcal{A}$ and let

$$1 \xrightarrow{a_1} a_1 \xrightarrow{a_2} a_2 \quad \cdots \quad a_{n-1} \xrightarrow{a_n} a_n$$

be a successful path with label u. Then the state a_n is final and hence $a_n \in S$. Similarly, since $1 \xrightarrow{a_1} a_1$ is a transition, one has necessarily $a_1 \in P$. Finally, since for $1 \leqslant i \leqslant n-1$, $a_i \xrightarrow{a_{i+1}} a_{i+1}$ is a transition, the word a_ia_{i+1} is not in N. Consequently, u belongs to L.

Conversely, if $u = a_1 \cdots a_n \in L$, one has $a_1 \in P$, $a_n \in S$ and, for $1 \leqslant i \leqslant n$, $a_ia_{i+1} \notin N$. Thus $1 \xrightarrow{a_1} a_1 \xrightarrow{a_2} a_2 \quad \cdots \quad a_{n-1} \xrightarrow{a_n} a_n$ is a successful path of $\mathcal{A}$ and $\mathcal{A}$ accepts u. Therefore, the language accepted by $\mathcal{A}$ is L. □

For a local language containing the empty word, the previous construction can be easily modified by taking $S \cup \{1\}$ as the set of final states.

Example 5.1. Let $A = \{a, b, c\}$, $P = \{a, b\}$, $S = \{a, c\}$ and $N = \{ab, bc, ca\}$. Then the automaton in Figure 5.1 recognises the language $L = (PA^* \cap A^*S) - A^*NA^*$.

[4] P stands for prefix, S for suffix and N for non-factor.

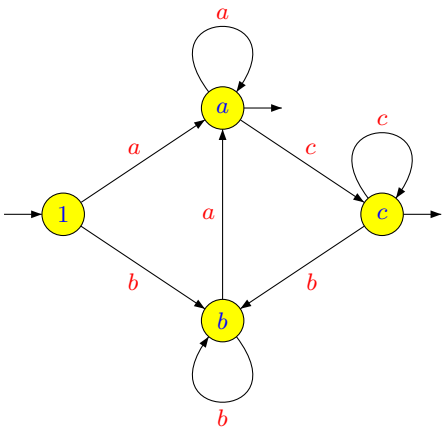


Figure 5.1. An automaton recognising a local language.

Note also that the automaton $\mathcal{A}$ described in Proposition 5.21 has a special property: all the transitions with label a have the same end, namely the state a. More generally, we shall say that a deterministic automaton (not necessarily complete) $\mathcal{A} = (Q, A, \cdot\,)$ is *local* if, for each letter a, the set $\{q \cdot a \mid q \in Q\}$ contains at most one element. Local languages have the following characterisation:

Proposition 5.22. *A rational language is local if and only if it is recognised by a local automaton.*

Proof. One direction follows from Proposition 5.21. To prove the opposite direction, consider a local automaton $\mathcal{A} = (Q, A, \cdot, q_0, F)$ recognising a language L and let

$$\begin{aligned} P &= \{a \in A \mid q_0 \cdot a \text{ is defined } \}, \\ S &= \{a \in A \mid \text{ there exists } q \in Q \text{ such that } q \cdot a \in F\}, \\ N &= \{x \in A^2 \mid x \text{ is the label of no path in } \mathcal{A} \,\} \\ K &= (PA^* \cap A^*S) - A^*NA^*. \end{aligned}$$

Let $u = a_1 \cdots a_n$ be a nonempty word of L and let $q_0 \xrightarrow{a_1} q_1 \cdots q_{n-1} \xrightarrow{a_n} q_n$ be a successful path with label u. Necessarily, $a_1 \in P$, $a_n \in S$ and, for $1 \leqslant i \leqslant n-1$, $a_i a_{i+1} \notin N$. Consequently, $u \in K$, which shows that $L - 1$ is contained in K.

Let now $u = a_1 \cdots a_n$ be a nonempty word of K. Then $a_1 \in P$, $a_n \in S$, and, for $1 \leqslant i \leqslant n-1$, $a_i a_{i+1} \notin N$. Since $a_1 \in P$, the state $q_1 = q_0 \cdot a_1$ is well defined. Furthermore, since $a_1 a_2 \notin N$, $a_1 a_2$ is the label of some path $p_0 \xrightarrow{a_1} p_1 \xrightarrow{a_2} p_2$ in $\mathcal{A}$. But since $\mathcal{A}$ is a local automaton, $q_0 \cdot a_1 = p_0 \cdot a_1$. It follows that the word $a_1 a_2$ is also the label of the path $q_0 \xrightarrow{a_1} p_1 \xrightarrow{a_2} p_2$. One can show in the same way by induction that there exists a sequence of states p_i $(0 \leqslant i \leqslant n)$ such that $a_i a_{i+1}$ is the label of a path $p_{i-1} \xrightarrow{a_i} p_i \xrightarrow{a_{i+1}} p_{i+1}$ of $\mathcal{A}$. Finally, since $a_n \in S$, there is a state q such that $q \cdot a_n \in F$. But since $\mathcal{A}$ is a local automaton, one has $q \cdot a_n = p_{n-1} \cdot a_n = p_n$, whence $p_n \in F$. Therefore $q_0 \xrightarrow{a_1} p_1 \cdots p_{n-1} \xrightarrow{a_n} p_n$ is a successful path in $\mathcal{A}$ and its label u is accepted by $\mathcal{A}$. Thus $K = L - 1$. □

Local languages are stable under various operations:

Proposition 5.23. *Let A_1 and A_2 be two disjoint subsets of the alphabet A and let $L_1 \subseteq A_1^*$ and $L_2 \subseteq A_2^*$ be two local languages. Then the languages $L_1 + L_2$ and L_1L_2 are local languages.*

Proof. Let $\mathcal{A}_1$ [$\mathcal{A}_2$] be a local automaton recognising L_1 [L_2]. The proofs of Propositions 4.9 and 4.12 give an automaton recognising $L_1 + L_2$ and L_1L_2. A simple verification shows that these constructions produce a local automaton when $\mathcal{A}_1$ and $\mathcal{A}_2$ are local. □

Proposition 5.24. *Let L be a local language. Then the language L^* is a local language.*

Proof. Let $\mathcal{A}$ be a local automaton recognising L. The proof of Proposition 4.13 gives an automaton recognising L^*. A simple verification shows that this construction produces a local automaton when $\mathcal{A}$ is local. □

5.2 Glushkov's algorithm

Glushkov's algorithm is an efficient way to convert a rational expression to a nondeterministic automaton.

A rational expression is said to be *linear* if each letter has at most one occurrence in the expression. For instance, the expression

$$[a_1a_2(a_3a_4)^* + (a_5a_6)^*a_7]^* \tag{5.1}$$

is linear. One can linearise a rational expression by replacing each occurrence of a letter by a distinct symbol. For instance, the expression (5.1) is a linearisation of the expression $e = [ab(ba)^* + (ac)^*b]^*$. Now, given an automaton for the linearisation e' of e, it is easy to obtain an automaton for e, simply by replacing the letters of e' by the corresponding letters in e. For instance, starting from the automaton $\mathcal{A}$ which recognises $[(a_1a_2)^*a_3]^*$, one gets a nondeterministic automaton $\mathcal{A}'$ which recognises $[(ab)^*a]^*$ by replacing a_1 and a_3 by a and a_2 by b, as shown in Figure 5.2.

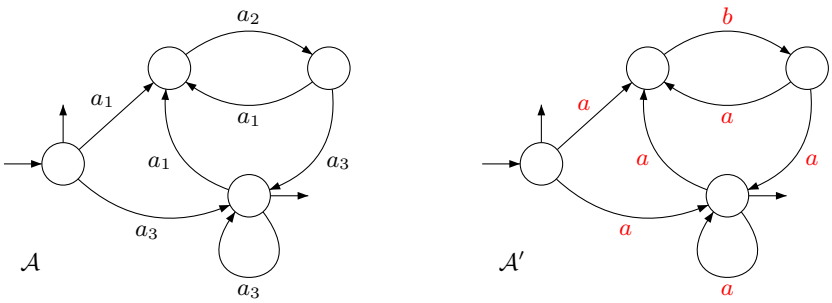


Figure 5.2. Construction of an automaton recognising $[(ab)^*a]^*$.

It remains to find an algorithm to compute the automaton of a linear expression.

Proposition 5.25. *Every linear expression represents a local language.*

Proof. The proof works by induction on the formation rules of a linear expression. First, the languages represented by 0, 1 and a, for $a \in A$, are local languages. Next, by Proposition 5.24, if e represents a local language, then so does e^*. Let now e and e' be two linear expressions and suppose that the expression $(e+e')$ is still linear. Let B $[B']$ be the set of letters occurring in e $[e']$. Since $(e+e')$ is linear, the letters of B $[B']$ do not occur in e' $[e]$. In other words, B and B' are disjoint and the local language represented by e $[e']$ is contained in B^* $[B'^*]$. By Proposition 5.23, the language represented by $(e+e')$ is also a local language. A similar argument applies for the language represented by ee'. □

Proposition 5.21 allows one to compute a deterministic automaton recognising a local language. It suffices to test whether the empty word belongs to L and to compute the sets

$$\begin{aligned} P(L) &= \{a \in A \mid aA^* \cap L \neq \emptyset\},\\ S(L) &= \{a \in A \mid A^*a \cap L \neq \emptyset\},\\ F(L) &= \{x \in A^2 \mid A^*xA^* \cap L \neq \emptyset\}. \end{aligned}$$

This can be done by recursion, given a linear rational expression representing the language. We first execute the procedure

EmptyWord(e: linear expression): **boolean**;

which determines whether the empty word belongs to the language represented by e.

EmptyWord(0) = **false**;
EmptyWord(1) = **true**;
EmptyWord(a) = **false** for all $a \in A$;
EmptyWord($e + e'$) = EmptyWord(e) **or** EmptyWord(e');
EmptyWord($e\cdot e'$) = EmptyWord(e) **and** EmptyWord(e');
EmptyWord(e^*) = **true**;

Now P, S and F are computed by the following recursive procedures:

P(0) = ∅;
P(1) = ∅;
P(a) = {a} for all $a \in A$;
P($e + e'$) = P(e) ∪ P(e');
if EmptyWord(e)
  then P($e\cdot e'$) = P(e) ∪ P(e')
  else $P(e\cdot e') = P(e)$;
P(e^*) = P(e);

S(0) = ∅;
S(1) = ∅;
S(a) = {a} for all $a \in A$;
S($e + e'$) = S(e) ∪ S(e');
if EmptyWord(e')
  then S($e\cdot e'$) = S(e) ∪ S(e')
  else S($e\cdot e'$) = S(e');
S(e^*) = S(e);

F(0) = ∅;
F(1) = ∅;
F(a) = ∅ for all $a \in A$;
F($e + e'$) = F(e) ∪ F(e');
F($e\cdot e'$) = F(e) ∪ F(e') ∪ S(e)P(e');
F(e^*) = F(e) ∪ S(e)P(e);

In summary, Glushkov's algorithm to convert a rational expression e to a nondeterministic automaton works as follows:

(1) Linearise e to e' and memorise the coding of the letters.
(2) Compute recursively the sets $P(e')$, $S(e')$ and $F(e')$. Then compute a deterministic automaton $\mathcal{A}'$ recognising e'.
(3) Convert $\mathcal{A}'$ to a nondeterministic automaton $\mathcal{A}$ recognising e.

Example 5.2. Consider the rational expression $e = (a(ab)^*)^* + (ba)^*$. We first linearise e to $e' = (a_1(a_2a_3)^*)^* + (a_4a_5)^*$. Let $L = L(e)$ and $L' = L(e')$. To compute the sets P, S and F, one can either use the above-mentioned recursive procedures, or proceed to a direct computation (this method is usually preferred in a computation by hand...). Recall that P [S] is the set of first [last] letters of the words of L'. We get

$$P = \{a_1, a_4\} \quad \text{and} \quad S = \{a_1, a_3, a_5\}$$

Note that a_1 belongs to S since a_1 is a word of L'.

Next we compute the set F of all words of length 2 that are factors of some word of L'. We get $F = \{a_1a_2, a_1a_1, a_2a_3, a_3a_1, a_3a_2, a_4a_5, a_5a_4\}$. For instance, a_3a_1 is a factor of $a_1a_2a_3a_1$ and a_3a_2 is a factor of $a_1a_2a_3a_2a_3$. Since the empty word belongs to L', the state 1 is final and we finally obtain the automaton represented in Figure 5.3. Since this automaton is local, there is actually no need to write the labels on the transitions. We now convert this automaton to a nondeterministic automaton recognising L, represented in Figure 5.4.

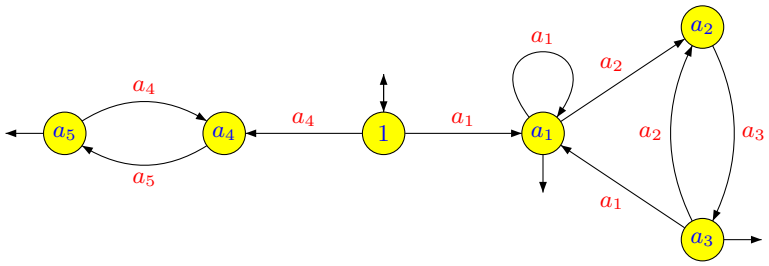


Figure 5.3. A local automaton recognising L'.

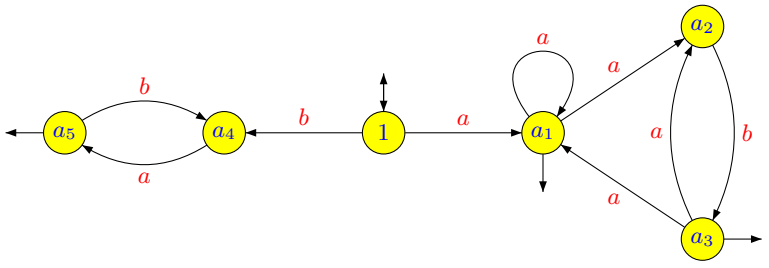


Figure 5.4. A nondeterministic automaton recognising L.

To get a deterministic automaton, it remains to apply the algorithm described in Section 3.2, which leads to the automaton represented in Figure 5.5.

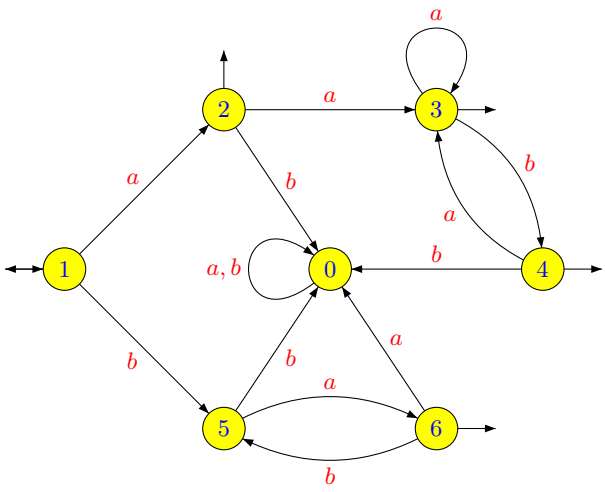


Figure 5.5. A deterministic automaton recognising L.

5.3 Linear equations

In this section, we give an algorithm to convert an automaton to a rational expression. The algorithm amounts to solving a system of linear equations on languages. We first consider an equation of the form

$$X = KX + L, \tag{5.2}$$

where K and L are languages and X is the unknown. When K does not contain the empty word, the equation admits a unique solution.

Proposition 5.26 (Arden's Lemma)**.** *If K does not contain the empty word, then $X = K^*L$ is the unique solution of the equation $X = KX + L$.*

Proof. Replacing X by K^*L in the expression $KX + L$, one gets

$$K(K^*L) + L = K^+L + L = (K^+ + 1)L = K^*L,$$

and hence $X = K^*L$ is a solution of (5.2). To prove uniqueness, consider two solutions X_1 and X_2 of (5.2). By symmetry, it suffices to show that each word u of X_1 also belongs to X_2. Let us prove this result by induction on the length of u.

If $|u| = 0$, u is the empty word and if $u \in X_1 = KX_1 + L$, then necessarily $u \in L$ since $1 \notin K$. But in this case, $u \in KX_2 + L = X_2$. For the induction step, consider a word u of X_1 of length $n+1$. Since $X_1 = KX_1 + L$, u belongs either to L or to KX_1. If $u \in L$, then $u \in KX_2 + L = X_2$. If $u \in KX_1$ then $u = kx$ for some $k \in K$ and $x \in X_1$. Since k is not the empty word, one has necessarily $|x| \leqslant n$ and hence by induction $x \in X_2$. It follows that $u \in KX_2$ and finally $u \in X_2$. This concludes the induction and the proof of the proposition. □

If K contains the empty word, uniqueness is lost.

Proposition 5.27. *If K contains the empty word, the solutions of (5.2) are the languages of the form K^*M with $L \subseteq M$.*

Proof. Since K contains the empty word, one has $K^+ = K^*$. If $L \subseteq M$, one has $L \subseteq M \subseteq K^*M$. It follows that the language K^*M is solution of (5.2) since

$$K(K^*M) + L = K^+M + L = K^*M + L = K^*M.$$

Conversely, let X be a solution of (5.2). Then $L \subseteq X$ and $KX \subseteq X$. Consequently, $K^2X \subseteq KX \subseteq X$ and by induction, $K^nX \subseteq X$ for all n. It follows that $K^*X = \sum_{n \geqslant 0} K^nX \subseteq X$. The language X can thus be written as K^*M with $L \subseteq M$: it suffices to take $M = X$. □

In particular, if K contains the empty word, then A^* is the maximal solution of (5.2) and the minimal solution is K^*L.

Consider now a system of the form

$$\begin{aligned} X_1 &= K_{1,1}X_1 + K_{1,2}X_2 + \cdots + K_{1,n}X_n + L_1 \\ X_2 &= K_{2,1}X_1 + K_{2,2}X_2 + \cdots + K_{2,n}X_n + L_2 \\ &\vdots \quad \vdots \\ X_n &= K_{n,1}X_1 + K_{n,2}X_2 + \cdots + K_{n,n}X_n + L_n \end{aligned} \tag{5.3}$$

We shall only consider the case when the system admits a unique solution.

Proposition 5.28. *If, for $1 \leqslant i, j \leqslant n$, the languages $K_{i,j}$ do not contain the empty word, the system (5.3) admits a unique solution. If further the $K_{i,j}$ and the L_i are rational languages, then the solutions X_i of (5.3) are rational languages.*

Proof. The case $n = 1$ is handled by Proposition 5.26. Suppose that $n > 1$. Consider the last equation of the system (5.3), which can be written

$$X_n = K_{n,n}X_n + (K_{n,1}X_1 + \cdots + K_{n,n-1}X_{n-1} + L_n)$$

According to Proposition 5.26, the unique solution of this equation is

$$X_n = K_{n,n}^*(K_{n,1}X_1 + \cdots + K_{n,n-1}X_{n-1} + L_n)$$

Replacing X_n by this expression in the $n-1$ first equations, we obtain a system of $n-1$ equations with $n-1$ unknowns and one can conclude by induction. □

We shall now associate a system of linear equations with every finite automaton $\mathcal{A} = (Q, A, E, I, F)$. Let us set, for $p, q \in Q$,

$$K_{p,q} = \{a \in A \mid (p, a, q) \in E\}$$
$$L_q = \begin{cases} 1 & \text{if } q \in F \\ 0 & \text{if } q \notin F \end{cases}$$

The solutions of the system defined by these parameters are the languages recognised by the automata

$$\mathcal{A}_q = (Q, A, E, \{q\}, F)$$

More precisely, we get the following result:

Proposition 5.29. *The system (5.3) admits a unique solution $(R_q)_{q\in Q}$, given by the formula*

$$R_q = \{u \in A^* \mid \text{ there is a path with label } u \text{ from } q \text{ to } F\}$$

Moreover, the language recognised by $\mathcal{A}$ is $\sum_{q\in I} R_q$.

Proof. Since the languages $K_{p,q}$ do not contain the empty word, Proposition 5.28 shows that the system (5.3) admits a unique solution. It remains to verify that the family $(R_q)_{q\in Q}$ is a solution of the system, that is, it satisfies for all $q \in Q$ the formula

$$R_q = K_{q,1}R_1 + K_{q,2}R_2 + \cdots + K_{q,n}R_n + L_q \tag{5.4}$$

Let S_q denote the right-hand side of (5.4). If $u \in R_q$, then u is by definition the label of a path from q to a final state f. If u is the empty word, one has necessarily $q = f$ and hence $L_q = 1$. Thus $u \in S_q$ in this case. Otherwise, let (q, a, q') be the first transition of the path. One has $u = au'$, where u' is the label of a path from q' to f. Then one has $a \in K_{q,q'}$, $u' \in R_{q'}$ and finally $u \in S_q$.

Conversely, let $u \in S_q$. If $u = 1$, one has necessarily $u \in L_q$, whence $q \in F$ and $u \in R_q$. Otherwise there is a state q' such that $u \in K_{q,q'}R_{q'}$. Therefore, $u = au'$ for some $a \in K_{q,q'}$ and $u' \in R_{q'}$. On the one hand, (q, a, q') is a transition of $\mathcal{A}$ by definition of $K_{q,q'}$ and on the other hand u' is the label of a final path starting in q'. The composition of these paths gives a final path with label u starting in q. Therefore $u \in R_q$ and thus $R_q = S_q$. □

Example 5.3. For the automaton represented in Figure 5.6,

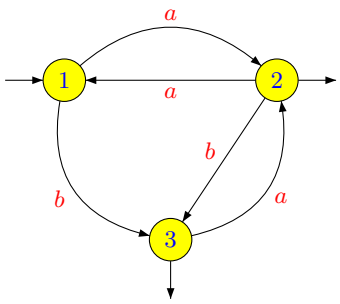


Figure 5.6. An automaton.

the system can be written

$$\begin{aligned} X_1 &= aX_2 + bX_3 \\ X_2 &= aX_1 + bX_3 + 1 \\ X_3 &= aX_2 + 1 \end{aligned}$$

Substituting $aX_2 + 1$ for X_3, one gets the equivalent system

$$X_1 = aX_2 + b(aX_2 + 1) = (a + ba)X_2 + b$$

$$X_2 = aX_1 + b(aX_2 + 1) + 1 = aX_1 + baX_2 + (b + 1)$$
$$X_3 = aX_2 + 1$$

Substituting $(a + ba)X_2 + b$ for X_1, one gets a third equivalent system

$$X_1 = (a + ba)X_2 + b$$
$$X_2 = a((a + ba)X_2 + b) + baX_2 + (b + 1) = (aa + aba + ba)X_2 + (ab + b + 1)$$
$$X_3 = aX_2 + 1$$

The solution of the second equation is

$$X_2 = (aa + aba + ba)^*(ab + b + 1)$$

Replacing X_2 by its value in the two other equations, we obtain

$X_1 = (a+ba)(aa+aba+ba)^*(ab+b+1)+b$ and $X_3 = a(aa+aba+ba)^*(ab+b+1)+1$

Finally, the language recognised by the automaton is

$$X_1 = (a + ba)(aa + aba + ba)^*(ab + b + 1) + b$$

since 1 is the unique initial state.

5.4 Extended automata

The use of equations is not limited to deterministic automata. The same technique applies to nondeterministic automata and to more powerful automata, in which the transition labels are not letters, but rational languages.

An *extended automaton* is a quintuple $\mathcal{A} = (Q, A, E, I, F)$, where Q is a set of states, A is an alphabet, E is a subset of $Q \times \mathrm{Rat}(A^*) \times Q$, called the set of transitions, I [F] is the set of initial [final] states. The label of a path

$$c = (q_0, L_1, q_1), (q_1, L_2, q_2), \ldots, (q_{n-1}, L_n, q_n)$$

is the rational language $L_1L_2 \cdots L_n$. The definition of a successful path is unchanged. A word is accepted by $\mathcal{A}$ if *it belongs to* the label of a successful path.

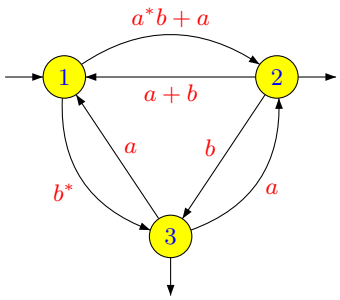


Figure 5.7. An extended automaton.

In the example represented in Figure 5.7, the set of transitions is

$$\{(1, a^*b + a, 2), (1, b^*, 3), (2, a + b, 1), (2, b, 3), (3, a, 1), (3, a, 2)\}$$

Let $\mathcal{A} = (Q, A, E, I, F)$ be an extended automaton. For all $p, q \in Q$, we let $K_{p,q}$ denote the label of the transition from p to q. Notice that $K_{p,q}$ might possibly be the empty language. We also put

$$L_q = \begin{cases} 1 & \text{if there is a path labelled by 1 from } q \text{ to } F \\ 0 & \text{otherwise} \end{cases}$$

Yet the associated system does not necessarily fulfil the condition $1 \notin K_{i,j}$ and Proposition 5.29 needs to be modified as follows:

Proposition 5.30. *The system (5.3) has a minimal solution $(R_q)_{q\in Q}$, given by the formula*

$$R_q = \{u \in A^* \mid \text{ there is a path labelled by } u \text{ from } q \text{ to } F\}$$

In particular the language recognised by $\mathcal{A}$ is $\sum_{q\in I} R_q$.

Proof. Let us first verify that the family $(R_q)_{q\in Q}$ is indeed a solution of (5.3), i.e. it satisfies, for all $q \in Q$:

$$R_q = K_{q,1}R_1 + K_{q,2}R_2 + \cdots + K_{q,n}R_n + L_q \tag{5.5}$$

Denote by S_q the right-hand side of (5.5). If $u \in R_q$, then u is by definition the label of a path from q to F. If $u = 1$, one has $L_q = 1$ and thus $u \in S_q$. Otherwise, let (q, u_1, q') be the first transition of the path. One has $u = u_1u'$, where u' is the label of a path from q' to F. Therefore $u_1 \in K_{q,q'}$, $u' \in R_{q'}$ and finally $u \in S_q$.

Conversely, let $u \in S_q$. If $u = 1$, one has necessarily $u \in L_q$, whence $q \in F$ and $u \in R_q$. Otherwise, there is a state q' such that $u \in K_{q,q'}R_{q'}$. Thus $u = u_1u'$ for some $u_1 \in K_{q,q'}$ and $u' \in R_{q'}$. On the one hand, (q, u_1, q') is a transition of $\mathcal{A}$ by the definition of $K_{q,q'}$ and on the other hand, u' is the label of a path from q' to F. Therefore $u = u_1u'$ is the label of a path from q to F and $u \in R_q$. Consequently $R_q = S_q$.

It remains to verify that if $(X_q)_{q\in Q}$ is a solution of the system, then $R_q \subseteq X_q$ for all $q \in Q$. If $u \in R_q$, there exists a path labelled by u from q to F:

$$(q_0, u_1, q_1)(q_1, u_2, q_2)\cdots(q_{r-1}, u_r, q_r)$$

with $q_0 = q$, $q_r \in F$, $u_i \in K_{q_{i-1},q_i}$ and $u_1u_2\cdots u_r = u$. Let us show by induction on $r - i$ that $u_{i+1}\cdots u_r$ belongs to X_{q_i}. By hypothesis, the X_q are solutions of

$$X_q = K_{q,1}X_1 + K_{q,2}X_2 + \cdots + K_{q,n}X_n + L_q$$

In particular, since $q_r \in F$, one has $1 \in L_{q_r}$ and hence $1 \in X_{q_r}$, which gives the result for $r - i = 0$. Furthermore, if $u_{i+1}\cdots u_r$ is an element of X_{q_i}, the inclusion $K_{q_{i-1},q_i}X_{q_i} \subseteq X_{q_{i-1}}$ shows that $u_iu_{i+1}\cdots u_r$ is an element of $X_{q_{i-1}}$, which concludes the induction. In particular, $u = u_1\cdots u_r \in X_q$. □

Example 5.4. For the extended automaton represented in Figure 5.7, the system can be written

$$X_1 = (a^*b + a)X_2 + b^*X_3$$
$$X_2 = (a + b)X_1 + bX_3 + 1$$

$$X_3 = aX_1 + aX_2 + 1$$

Replacing X_3 by $aX_1 + aX_2 + 1$, and observing that $a + b^*a = b^*a$, we obtain the equivalent system

$$\begin{aligned}
X_1 &= (a^*b + a)X_2 + b^*(aX_1 + aX_2 + 1) = b^*aX_1 + (a^*b + b^*a)X_2 + b^* \\
X_2 &= (a + b)X_1 + b(aX_1 + aX_2 + 1) + 1 = (a + b + ba)X_1 + baX_2 + b + 1 \\
X_3 &= aX_1 + aX_2 + 1
\end{aligned}$$

We deduce from the second equation

$$X_2 = (ba)^*((a + b + ba)X_1 + b + 1)$$

and replacing X_2 by its value in the first equation, we obtain

$$\begin{aligned}
X_1 &= b^*aX_1 + (a^*b + b^*a)(ba)^*((a + b + ba)X_1 + b + 1) + b^* \\
&= (b^*a + (a^*b + b^*a)(ba)^*(a + b + ba))X_1 + (a^*b + b^*a)(ba)^*(b + 1) + b^*
\end{aligned}$$

Finally, the language recognised by the automaton is

$$X_1 = \big(b^*a + (a^*b + b^*a)(ba)^*(a + b + ba)\big)^*[(a^*b + b^*a)(ba)^*(b + 1) + b^*]$$

since 1 is the unique initial state.

5.5 Kleene's theorem

We are now ready to state the most important result of automata theory.

Theorem 5.31 (Kleene (1954))**.** *A language is rational if and only if it is recognisable.*

Proof. It follows from Proposition 5.29 that every recognisable language is rational. Corollary 4.10 states that every finite language is recognisable. Furthermore, Propositions 4.9, 4.12 and 4.13 show that recognisable languages are closed under union, product and star. Thus every rational language is recognisable. □

The following corollary is now a consequence of Propositions 2.3, 4.9, 4.8, 4.11, 4.12, 4.13, 4.15 and 4.16.

Corollary 5.32. *Recognisable* [*rational*] *languages are closed under Boolean operations, product, star, quotients, morphisms and inverses of morphisms.*

We conclude this section by proving some elementary decidability results on recognisable languages. Recall that a property is *decidable* if there is an algorithm to check whether this property holds or not. We shall also often use the expressions "given a recognisable language L" or "given a rational language L". As long as only decidability is concerned, it makes no difference to give a language by a nondeterministic automaton, a deterministic automaton or a regular expression, since there are algorithms to convert one of the forms to the other. However, the chosen representation is important for complexity issues, which will not be discussed here.

Theorem 5.33. *Given a recognisable language L, the following properties are decidable:*

(1) *whether a given word belongs to L,*
(2) *whether L is empty,*
(3) *whether L is finite,*
(4) *whether L is infinite,*

Proof. We may assume that L is given by a trimmed deterministic automaton $\mathcal{A} = (Q, A, \cdot, q_-, F)$.

(1) To test whether $u \in L$, it suffices to compute $q_- \cdot u$. If $q_- \cdot u \in F$, then $u \in L$; if $q_- \cdot u \notin F$, or if $q_- \cdot u$ is undefined, then $u \notin L$.

(2) Let us show that L is empty if and only if $F = \emptyset$. The condition $F = \emptyset$ is clearly sufficient. Since $\mathcal{A}$ is trimmed, every state of $\mathcal{A}$ is accessible. Now, if $\mathcal{A}$ has at least one final state q, there is a word u such that $q_- \cdot u = q$. Therefore $u \in L$ and L is nonempty.

(3) and (4). Let us show that L is finite if and only if $\mathcal{A}$ does not contain any loop. If $\mathcal{A}$ contains a loop $q \xrightarrow{u} q$, then L is infinite: indeed, since $\mathcal{A}$ is trimmed, there exist paths $q_- \xrightarrow{x} q$ and $q \xrightarrow{y} f$, where f is a final state and thus L contains all the words $xu^n y$. Conversely, if L is infinite, the proof of the pumping lemma shows that $\mathcal{A}$ contains a loop. Now, checking whether an automaton contains a loop is easy. Consider the directed graph G obtained from $\mathcal{A}$ by removing all the labels. Then $\mathcal{A}$ is loop-free if and only if G is acyclic, a property that can be checked by standard algorithms. One can for instance compute the transitive closure G' of G and check whether G' contains an edge of the form (q, q). □

We leave as an exercise to the reader to prove that the inclusion problem and the equality problem are decidable for two given recognisable languages.

6 Exercises

Section 1

Exercise 1. Let us say that two words x and y are *powers of the same word* if there exists a word z and two nonnegative integers n and m such that $x = z^n$ and $y = z^m$. Show that two words commute if and only if they are powers of the same word.

Exercise 2. Two words x and y are *conjugate* if there exist two words u and v such that $x = uv$ and $y = vu$.

(1) Show that two words are conjugate if and only if there exists a word z such that $xz = zy$.
(2) Conclude that the conjugacy relation is an equivalence relation.

Exercise 3. A word u is a *subword* of v if v can be written as

$$v = v_0 u_1 v_1 u_2 v_2 \cdots u_k v_k$$

where u_i and v_i are words (possibly empty) such that $u_1 u_2 \cdots u_k = u$. For instance, the words *baba* and *acab* are subwords of *abcacbab*.

(1) Show that the subword relation is a partial ordering on A^*.
(2) (Difficult) Prove that if A is finite, any infinite set of words contains two words, one of which is a subword of the other.

Section 2

Exercise 4. Simplify the following rational expressions
(1) $((abb)^*)^*$,
(2) $a^*b + a^*ba^*$,
(3) 0^*ab1^*,
(4) $(a^*b)^*a^*$.

Exercise 5. Let e, e_1, e_2, e_3 be four rational expressions. Verify that the following pairs of rational expressions represent the same language:
(1) $e_1 + e_2$ and $e_2 + e_1$,
(2) $((e_1 + e_2) + e_3)$ and $(e_1 + (e_2 + e_3))$,
(3) $((e_1e_2)e_3)$ and $(e_1(e_2e_3))$,
(4) $(e_1(e_2 + e_3))$ and $((e_1e_2) + (e_1e_3))$,
(5) $((e_1 + e_2)e_3)$ and $((e_1e_3) + (e_2e_3))$,
(6) e^{**} and e^*,
(7) $0 + e$ and e,
(8) $1e$ and e,
(9) 0^* and 1,
(10) $(e_1 + e_2)^*$ and $(e_1^*e_2)^*e_1^*$.

Section 3

Exercise 6. A well-parenthesised word is a word on the alphabet $\{(,)\}$ containing as many left parenthesises as right parenthesises and such that each of its prefixes contains at least as many left parenthesises as right parenthesises. For instance, $(((())())())$ is a well-parenthesised word, but $(((())()))())$ is not.

Prove that the language of all well parenthesised words is not recognisable.

Exercise 7. Prove that the following languages are not recognisable:
(1) $\{u \in \{a,b\}^* \mid |u|_a = |u|_b\}$,
(2) $\{a^nb^m \mid n, m \geqslant 0 \text{ and } n = 2m\}$,
(3) $\{u \in A^* \mid u = \tilde{u}\}$
(4) $\{v\tilde{v}w \mid v, w \in \{a,b\}^+\}$
(5) $\{u \in a^* \mid |u| \text{ is a prime number }\}$

Exercise 8. Let $A = \{a,b,c\}$. Show that the language

$$L = \{(ab)^nc^n \mid n > 0\} + A^*bbA^* + A^*aaA^*$$

is not recognisable, but satisfies the pumping lemma: there is an integer $n > 0$ such that every word u of L of length greater than or equal to n can be factored as $u = xyz$ with $x, y, z \in A^*$, $|xy| \leqslant n$, $y \neq 1$ and, for all $k \geqslant 0$, $xy^kz \in L$.

Exercise 9. Prove that, for each $n > 0$, the language

$$L_n = \{u \in \{a,b\}^* \mid |u|_a \equiv |u|_b \bmod n\}$$

is recognisable, and compute its minimal automaton.

Exercise 10. Let $\mathcal{A}_n = (\{0,1,\ldots,n-1\}, \{a,b\}, E_n, \{0\}, \{0\})$ be the nondeterministic automaton defined by

$$E_n = \{(i,a,i+1) \mid 0 \leqslant i \leqslant n-1\} \cup \{(n-1,a,0)\} \cup \{(i,b,i) \mid 1 \leqslant i \leqslant n-1\} \cup \{(i,b,0) \mid 1 \leqslant i \leqslant n-1\}$$

and represented in Figure 6.1. Show that any deterministic automaton equivalent to $\mathcal{A}_n$ has at least 2^n states.

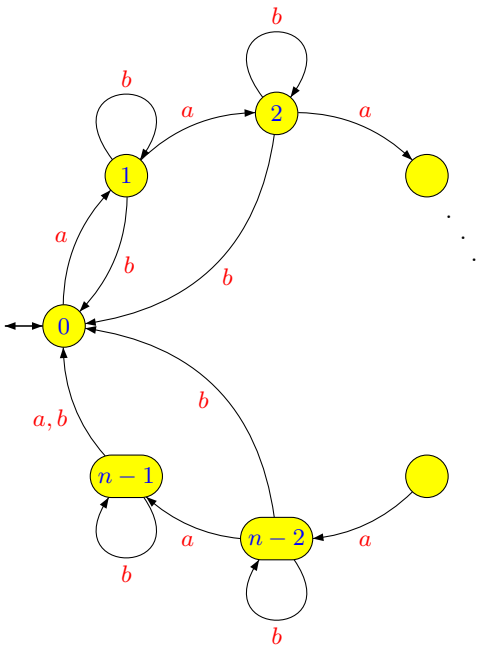


Figure 6.1. The automaton $\mathcal{A}_n$.

Section 4

Exercise 11. The *shuffle* of two words u and v is the language $u \sqcup v$ consisting of all words

$$u_1 v_1 u_2 v_2 \cdots u_k v_k$$

where $k \geqslant 0$, the u_i and v_i are words of A^*, such that $u_1 u_2 \cdots u_k = u$ and $v_1 v_2 \cdots v_k = v$. For instance,

$$ab \sqcup ba = \{abab, abba, baba, baab\}.$$

By extension, the shuffle of two languages K and L is the language

$$K \sqcup L = \bigcup_{u \in K, v \in L} u \sqcup v$$

Prove that the shuffle is a commutative and associative operation, which distributes over union. Show that if K and L are recognisable, then $K \sqcup L$ is recognisable.

Exercise 12. Compute the minimal automaton of the language $(a(ab)^*b)^*$.

Exercise 13. Consider the sequence of languages D_n defined by $D_0 = \{1\}$ and $D_{n+1} = (aD_nb)^*$. Compute the minimal automaton of D_0, D_1 and D_2. Guess from these examples the minimal automaton of D_n and prove that your guess is correct.

Let, for $u \in A^*$, $||u|| = |u|_a - |u|_b$. Show that, for all $n \geqslant 0$, the following conditions are equivalent:

(1) $u \in D_n$,

(2) $||u|| = 0$ and for all prefixes v of u, $0 \leqslant ||v|| \leqslant n$.

Conclude that, for all $p, q \geqslant 0$, $D_p \sqcup D_q = D_{p+q}$.

Let $D = \bigcup_{n \geqslant 0} D_n$. Show that that D is not recognisable.

Section 5

Exercise 14. For each of these languages on the alphabet $\{a, b\}$, compute their minimal automaton and a rational expression representing them:

(1) The language $a(a + b)^*a$.

(2) The set of words containing two consecutive a.

(3) The set of words with an even number of b and an odd number of a.

(4) The set of words not containing the factor $abab$.

Exercise 15. This exercise relies on the notion of a *countable set*. Let A be a finite nonempty alphabet. Prove that A^* and the set of recognisable languages of A^* are countable. (Hint: one can enumerate the finite automata on the alphabet A). Arguing on the fact that the set of subsets of an infinite set in uncountable, deduce that there exist some nonrecognisable languages on A^*.

Use a similar argument to prove that there are some nonrational languages on A^*.

7 Notes

The material of this chapter is quite standard. For a complete introduction to automata theory, the reader is referred to specialised books [27, 64, 138].

Chapter IV

Recognisable and rational sets

The notions of rational and recognisable sets are usually defined for finitely generated free monoids, under the common name of regular sets. Following Eilenberg [42], one can extend these notions to arbitrary monoids. The price to pay is that Kleene's theorem does not extend to arbitrary monoids. Although the classes of rational and recognisable sets coincide in finitely generated free monoids, they form in general incomparable classes.

1 Rational subsets of a monoid

Let M be a monoid. We have already seen that the set $\mathcal{P}(M)$ of subsets of M is a semiring with union as addition and product defined by the formula

$$XY = \{xy \mid x \in X \quad \text{and} \quad y \in Y\}$$

For this reason, we shall adopt the notation we already introduced for languages. Union is denoted by $+$, the empty set by 0 and each singleton $\{m\}$ by m. This notation has the advantage that the identity of $\mathcal{P}(M)$ is denoted by 1.

The powers of a subset X of M are defined by induction by setting $X^0 = 1$, $X^1 = X$ and $X^n = X^{n-1}X$ for all $n > 1$. The *star* and *plus* operations are defined as follows:

$$X^* = \sum_{n \geqslant 0} X^n = 1 + X + X^2 + X^3 + \cdots$$
$$X^+ = \sum_{n > 0} X^n = X + X^2 + X^3 + \cdots$$

Note that X^* $[X^+]$ is the submonoid [subsemigroup] of M generated by X. The set of rational subsets of a monoid M is the smallest set $\mathcal{F}$ of subsets of M satisfying the following conditions:

(1) $\mathcal{F}$ contains 0 and the singletons of $\mathcal{P}(M)$,

(2) $\mathcal{F}$ is closed under finite union, product and star (in other words, if $X, Y \in \mathcal{F}$, then $X + Y \in \mathcal{F}$, $XY \in \mathcal{F}$ and $X^* \in \mathcal{F}$).

Example 1.1. In a finite monoid, all subsets are rational.

Example 1.2. The rational subsets of $\mathbb{N}^k$ are the *semilinear* sets, which are finite unions of subsets of the form

$$\{v_0 + n_1 v_1 + \cdots + n_r v_r \mid n_1, \ldots, n_r \in \mathbb{N}\}$$

where $v_0, v_1, \ldots, v_r$ are k-tuples of $\mathbb{N}^k$.

The rational subsets are by construction closed under finite union, product and star. They are also stable under morphisms.

Proposition 1.1. *Let $\varphi : M \to N$ be a monoid morphism. If R is a rational subset of M, then $\varphi(R)$ is a rational subset of N. If further φ is surjective, then for each rational subset S of N, there exists a rational subset R of M such that $\varphi(R) = S$.*

Proof. Denote by $\mathcal{F}$ the set of subsets K of M such that $\varphi(K)$ is a rational subset of N. The set $\mathcal{F}$ contains the finite sets, since, if K is finite, $\varphi(K)$ is also finite and hence rational. Furthermore, $\mathcal{F}$ is stable under union: if K and K' are in $\mathcal{F}$, that is, if $\varphi(K)$ and $\varphi(K')$ are rational, then $\varphi(K) + \varphi(K') = \varphi(K + K')$ is rational, and hence $K + K'$ is in $\mathcal{F}$. The proof that KK' and K^* are in $\mathcal{F}$ is similar but rests on the formulas

$$\varphi(KK') = \varphi(K)\varphi(K') \text{ and } \varphi(K^*) = (\varphi(K))^*.$$

It follows that $\mathcal{F}$ contains the rational subsets of M. By the definition of $\mathcal{F}$, this means that if L is rational, so is $\varphi(L)$.

For the second part of the statement, assume that φ is surjective and consider the set $\mathcal{S}$ of subsets S of N such that $S = \varphi(R)$ for some rational subset R of M. First observe that $\emptyset \in \mathcal{S}$ since $\varphi(\emptyset) = \emptyset$. Since φ is surjective, every element n of N can be written as $\varphi(m)$ for some $m \in M$. Thus $\mathcal{S}$ contains the singletons. Moreover, the formula

$$\varphi(R)\varphi(R') = \varphi(RR') \qquad \varphi(R + R') = \varphi(R) + \varphi(R') \qquad \varphi(R^*) = (\varphi(R))^*$$

show that $\mathcal{S}$ is closed under union, product and star. Consequently, $\mathcal{S}$ contains the rational subsets of N, which concludes the proof. □

However, the rational subsets of a monoid are not necessarily closed under intersection, as shown by the following counterexample:

Example 1.3. Let $M = a^* \times \{b, c\}^*$. Consider the rational subsets

$$R = (a, b)^*(1, c)^* = \{(a^n, b^n c^m) \mid n, m \geqslant 0\}$$
$$S = (1, b)^*(a, c)^* = \{(a^n, b^m c^n) \mid n, m \geqslant 0\}$$

Their intersection is

$$R \cap S = \{(a^n, b^n c^n) \mid n \geqslant 0\}$$

Let π be the projection from M onto $\{b, c\}^*$. If $R \cap S$ was rational, the language $\pi(R \cap S) = \{b^n c^n \mid n \geqslant 0\}$ would also be rational by Proposition 1.1. But Corollary III.3.5 shows that this language is not rational.

It also follows that the complement of a rational subset is not necessarily rational. Otherwise, the rational subsets of a monoid would be closed under union and complement and hence under intersection. However, let us mention the following result, which is beyond the scope of this course..

Theorem 1.2. *The rational subsets of a commutative monoid are closed under Boolean operations.*

Rational subsets are closed under direct products, in the following sense:

Theorem 1.3. *Let R_1 $[R_2]$ be a rational subset of a monoid M_1 $[M_2]$. Then $R_1 \times R_2$ is a rational subset of $M_1 \times M_2$.*

Proof. Let $\pi_1 : M_1 \to M_1 \times M_2$ and $\pi_2 : M_2 \to M_1 \times M_2$ be the morphisms defined by $\pi_1(m) = (m, 1)$ and $\pi_2(m) = (1, m)$. Then we have

$$R_1 \times R_2 = (R_1 \times \{1\})(\{1\} \times R_2) = \pi_1(R_1)\pi_2(R_2)$$

which shows, by Proposition 1.1, that $R_1 \times R_2$ is rational. □

Recall that a monoid is *finitely generated* if it admits a finite set of generators.

Proposition 1.4. *Each rational subset of a monoid M is a rational subset of a finitely generated submonoid of M.*

Proof. Consider the set $\mathcal{R}$ of subsets R of M that are rational subsets of a finitely generated submonoid of M. It is clear that $\mathcal{R}$ contains the empty set and the singletons, since $\{m\}$ is a rational subset of m^*. If R and S are in $\mathcal{R}$, there exist some finite subsets F and G of M such that R is a rational subset of F^* and S is a rational subset of G^*. It follows that $R + S$ and RS are rational subsets of $(F+G)^*$, and R^* is a rational subset of F^*. Consequently, $R+S$, RS and R^* are also in $\mathcal{R}$, proving that $\mathcal{R}$ contains the rational subsets of M. □

2 Recognisable subsets of a monoid

Recognisable languages are usually defined in terms of deterministic automata, which is the best definition from an algorithmic point of view. However, to handle the fine structure of recognisable languages, it is often appropriate to use a more abstract definition, due to Rabin and Scott [131], that uses monoids in place of automata. Although this definition will be mainly used in the context of free monoids, it is as simple to give it in a more general setting.

2.1 Recognition by monoid morphisms

Let $\varphi : M \to N$ be a monoid morphism. A subset L of M is *recognised* by φ if there exists a subset P of N such that

$$L = \varphi^{-1}(P).$$

If φ is surjective, we say that φ *fully recognises* L. Note that in this case, the condition $L = \varphi^{-1}(P)$ implies $P = \varphi(L)$.

Proposition 2.5. *If $\varphi : M \to N$ recognises L, then the morphism from M onto $\varphi(M)$ induced by φ fully recognises L.*

Proof. Since φ recognises L, there exists a subset P of N such that $L = \varphi^{-1}(P)$. It follows that $L = \varphi^{-1}(P \cap \varphi(M))$ which proves the result. □

Let us say that a congruence $\sim$ on M *saturates* a subset L of M if the conditions $u \in L$ and $u \sim v$ imply $v \in L$. Let us start by an elementary, but useful observation:

Proposition 2.6. *Let $\varphi : M \to N$ be a monoid morphism and let L be a subset of M. The following conditions are equivalent:*

(1) *L is recognised by φ,*

(2) *L is saturated by $\sim_\varphi$,*

(3) *$\varphi^{-1}(\varphi(L)) = L$.*

Proof. (1) implies (2). If L is recognised by φ, then $L = \varphi^{-1}(P)$ for some subset P of N. Thus if $x \in L$ and $x \sim_\varphi y$, one has $\varphi(x) \in P$ and since $\varphi(x) = \varphi(y)$, $y \in \varphi^{-1}(P) = L$. Therefore L is saturated by $\sim_\varphi$.

(2) implies (3). Suppose that L is saturated by $\sim_\varphi$. If $x \in \varphi^{-1}(\varphi(L))$, there exists $y \in L$ such that $\varphi(x) = \varphi(y)$, that is, $x \sim_\varphi y$. It follows that $x \in L$, which proves the inclusion $\varphi^{-1}(\varphi(L)) \subseteq L$. The opposite inclusion is trivial.

(3) implies (1). Setting $P = \varphi(L)$, one has $\varphi^{-1}(P) = L$. Thus L is recognised by φ. □

By extension, one also says that a *monoid N [fully] recognises* a subset L of a monoid M if there exists a [surjective] monoid morphism $\varphi : M \to N$ that recognises L.

Example 2.1. Let $(T, \oplus)$ be the commutative monoid defined on $\{0, 1, 2\}$ by

$$x \oplus y = \min\{x + y, 2\}$$

and let φ be the surjective morphism from $(\mathbb{N}, +)$ onto T defined by $\varphi(0) = 0$, $\varphi(1) = 1$ and $\varphi(n) = 2$ for all $n \geqslant 2$. The subsets of $\mathbb{N}$ recognised by φ are $\varphi^{-1}(\emptyset) = \emptyset$, $\varphi^{-1}(0) = \{0\}$, $\varphi^{-1}(1) = \{1\}$, $\varphi^{-1}(2) = \{2, 3, \ldots\}$, $\varphi^{-1}(\{0, 1\}) = \{0, 1\}$, $\varphi^{-1}(\{0, 2\}) = \{0, 2, 3, 4, \ldots\}$, $\varphi^{-1}(\{1, 2\}) = \{1, 2, 3, 4, \ldots\}$ and $\varphi^{-1}(\{0, 1, 2\}) = \mathbb{N}$.

Example 2.2. Let $M = B_2^1 = \{1, a, b, ab, ba, 0\}$ be the multiplicative monoid defined by the relations $aba = a$, $bab = b$, $aa = bb = 0$. Let $A = \{a, b\}$ and let $\varphi : A^* \to M$ be the morphism defined by $\varphi(a) = a$ and $\varphi(b) = b$. One has $\varphi^{-1}(1) = \{1\}$, $\varphi^{-1}(a) = (ab)^*a$, $\varphi^{-1}(b) = (ba)^*b$, $\varphi^{-1}(ab) = (ab)^+$, $\varphi^{-1}(ba) = (ba)^+$ and $\varphi^{-1}(0) = A^*aaA^* + A^*bbA^*$.

Example 2.3. Recall that U_2 denotes the monoid $\{1, a_1, a_2\}$ defined by $a_1a_1 = a_2a_1 = a_1$ and $a_1a_2 = a_2a_2 = a_2$. Let $\pi : A^* \to U_2$ be a morphism and let $A_1 = \{a \in A \mid \pi(a) = a_1\}$, $A_2 = \{a \in A \mid \pi(a) = a_2\}$ and $B = \{a \in A \mid \pi(a) = 1\}$. Then $\pi^{-1}(1) = B^*$, $\pi^{-1}(a_1) = A^*A_1B^*$ and $\pi^{-1}(a_2) = A^*A_2B^*$.

We conclude this section by a very useful property of a morphism recognising a subset of a monoid.

Proposition 2.7. *Let M be a monoid, P a subset of M and $\varphi : M \to N$ be a morphism recognising P. Then for each subset R of M, one has*

(1) *$\varphi(R \cap P) = \varphi(R) \cap \varphi(P)$,*

(2) $\varphi(R \cup P) = \varphi(R) \cup \varphi(P)$,
(3) $\varphi(R - P) = \varphi(R) - \varphi(P)$.

Proof. (1) The inclusion $\varphi(P \cap R) \subseteq \varphi(P) \cap \varphi(R)$ is clear. To prove the opposite inclusion, consider an element s of $\varphi(P) \cap \varphi(R)$. Then one has $s = \varphi(r)$ for some $r \in R$. It follows that $r \in \varphi^{-1}(s)$, therefore $r \in \varphi^{-1}(\varphi(P))$ and finally $r \in P$. Thus $r \in P \cap R$ and $s \in \varphi(P \cap R)$, which proves (1).

(2) is trivial.

(3) Let $r \in R - P$. Then $\varphi(r) \in \varphi(R)$. Furthermore, if $\varphi(r) \in \varphi(P)$, then $r \in P$ since φ recognises P. Therefore $\varphi(R - P) \subseteq \varphi(R) - \varphi(P)$.

To establish the opposite inclusion, consider an element x of M such that $\varphi(x) \in \varphi(R) - \varphi(P)$. Then $\varphi(x) \in \varphi(R)$ and thus $\varphi(x) = \varphi(r)$ for some $r \in R$. If $r \in P$, then $\varphi(x) \in \varphi(P)$, a contradiction. Therefore $r \in R - P$ and $\varphi(x) \in \varphi(R - P)$. □

Corollary 2.8. *Let M be a monoid, P a subset of M and $\varphi : M \to N$ be a morphism recognising P. If $P = X_1 - X_2$ with $X_2 \subseteq X_1$, then $\varphi(P) = \varphi(X_1) - \varphi(X_2)$.*

Proof. When X_2 is a subset of X_1, the conditions $P = X_1 - X_2$ and $X_2 = X_1 - P$ are equivalent. Proposition 2.7 now shows that $\varphi(X_2) = \varphi(X_1) - \varphi(P)$, which in turn gives $\varphi(P) = \varphi(X_1) - \varphi(X_2)$. □

2.2 Operations on sets

Simple operations on sets have a natural algebraic counterpart. We now study successively complement, intersection, union, inverses of morphisms and left and right quotients.

Proposition 2.9. *Let L be a subset of the monoid M. If L is recognised by $\varphi : M \to N$, then $M - L$ is also recognised by φ.*

Proof. If $L = \varphi^{-1}(P)$ then, by Proposition I.1.18, $M - L = \varphi^{-1}(N - P)$. □

For $1 \leqslant i \leqslant n$, let $\varphi_i : M \to M_i$ be a surjective monoid morphism. The *product* of these morphisms is the surjective morphism

$$\varphi : M \to \mathrm{Im}(\varphi) \subseteq M_1 \times \cdots \times M_n$$

defined by $\varphi(x) = (\varphi_1(x), \ldots, \varphi_n(x))$.

Proposition 2.10. *Let $L_1, \ldots, L_n$ be subsets of M. If each L_i is recognised by φ_i, then the sets $\cap_{1 \leqslant i \leqslant n} L_i$ and $\cup_{1 \leqslant i \leqslant n} L_i$ are recognised by their product φ.*

Proof. Suppose that $L_i = \varphi_i^{-1}(P_i)$ for some subset P_i of M_i. The result follows immediately from the two formulas

$$\bigcap_{1 \leqslant i \leqslant n} L_i = \varphi^{-1}(P_1 \times \cdots \times P_n)$$

$$\bigcup_{1 \leqslant i \leqslant n} L_i = \varphi^{-1}\Big(\bigcup_{1 \leqslant i \leqslant n} M_1 \times \cdots \times M_{i-1} \times P_i \times M_{i+1} \times \cdots \times M_n \Big)$$

□

Proposition 2.11. *Let $\eta : R \to M$ and $\varphi : M \to N$ be two morphisms of monoids. If φ recognises a subset L of M, then $\varphi \circ \eta$ recognises $\eta^{-1}(L)$.*

Proof. Suppose that $L = \varphi^{-1}(P)$ for some subset P of N. Then $\eta^{-1}(L) = \eta^{-1}(\varphi^{-1}(P)) = (\varphi \circ \eta)^{-1}(P)$. Thus $\varphi \circ \eta$ recognises $\eta^{-1}(L)$. □

Recall that, for each subset X of M and for each element s of M, the *left* [*right*] *quotient* $s^{-1}X$ [Xs^{-1}] of X by s is defined as follows:

$$s^{-1}X = \{t \in M \mid st \in X\} \quad \text{and} \quad Xs^{-1} = \{t \in M \mid ts \in X\}$$

More generally, for any subset K of M, the left [right] quotient $K^{-1}X$ [XK^{-1}] of X by K is

$$K^{-1}X = \bigcup_{s \in K} s^{-1}X = \{t \in M \mid \text{ there exists } s \in K \text{ such that } st \in X\}$$
$$XK^{-1} = \bigcup_{s \in K} Xs^{-1} = \{t \in M \mid \text{ there exists } s \in K \text{ such that } ts \in X\}$$

Proposition 2.12. *Let $\varphi : M \to N$ be a morphism of monoids. If φ recognises a subset L of M, it also recognises $K^{-1}L$ and LK^{-1} for every subset K of M.*

Proof. Suppose that $L = \varphi^{-1}(P)$ for some subset P of N, and let $R = \varphi(K)$. We claim that $\varphi^{-1}(R^{-1}P) = K^{-1}L$. Indeed, one has the following sequence of equivalent statements:

$$\begin{aligned} m \in \varphi^{-1}(R^{-1}P) &\iff \varphi(m) \in R^{-1}P \\ &\iff \text{there exists } r \in R \text{ such that } r\varphi(m) \in P \\ &\iff \text{there exists } k \in K \text{ such that } \varphi(k)\varphi(m) \in P \\ &\iff \text{there exists } k \in K \text{ such that } km \in \varphi^{-1}(P) \\ &\iff \text{there exists } k \in K \text{ such that } km \in L \\ &\iff m \in K^{-1}L \end{aligned}$$

Thus φ recognises $K^{-1}L$. A similar proof works for LK^{-1}. □

2.3 Recognisable sets

A subset of a monoid is *recognizable* if it is recognised by a finite monoid. We let $\mathrm{Rec}(M)$ denote the set of recognisable subsets of M. We shall see shortly in Section 3 that if $M = A^*$, then $\mathrm{Rec}(A^*)$ is the set of recognisable languages, as defined in Section III.3.1. Thus our new definition is consistent with the old one.

Notice that Propositions 2.9, 2.10, 2.11 and 2.12 subsume Propositions III.4.9, III.4.8, III.4.11, III.4.15, III.4.16 with a shorter proof. One can summarise these results in the following statement:

Corollary 2.13. *For any monoid M, $\mathrm{Rec}(M)$ is closed under Boolean operations and left and right quotients. Furthermore, if $\varphi : N \to M$ is a morphism, $L \in \mathrm{Rec}(M)$ implies $\varphi^{-1}(L) \in \mathrm{Rec}(N)$.*

Although Kleene's theorem does not extend to arbitrary monoids, a weaker property holds for finitely generated monoids.

Theorem 2.14 (McKnight)**.** *Let M be a monoid. The following conditions are equivalent:*

(1) *M is finitely generated,*
(2) *every recognisable subset of M is rational,*
(3) *the set M is a rational subset of M.*

Proof. ((1)) implies ((2)). Let M be a finitely generated monoid. Then there exists a finite alphabet A and a surjective morphism π from A^* onto M. Let R be a recognisable subset of M. By Corollary 2.13, the set $\pi^{-1}(R)$ is recognised by a finite monoid. Proposition 3.20)[1] implies that $\pi^{-1}(R)$ is recognised by a finite deterministic automaton and hence is rational by Kleene's theorem. Now, since $R = \pi(\pi^{-1}(R))$, Proposition 1.1 shows that R is rational.

((2)) implies ((3)) is clear, since M is a recognisable subset of M.

((3)) implies ((1)) follows immediately from Proposition 1.4. □

Let us come back to arbitrary monoids. We have seen that the intersection of two rational subsets is not necessarily rational, but that the intersection of two recognisable sets is recognisable. What about the intersection of a rational subset and a recognisable subset?

Proposition 2.15. *The intersection of a rational subset and of a recognisable subset of a monoid is rational.*

Proof. Let R be a rational subset of a monoid M. By Proposition 1.4, R is a rational subset of a finitely generated submonoid N of M. Let ι denote the identity mapping from N to M. Since N is finitely generated, there exist a finite alphabet A and a surjective morphism $\pi : A^* \to N$. By the second part of Proposition 1.1, there exists a rational language K of A^* such that $\iota(\pi(K)) = R$. If S is a recognisable subset of M, then $S \cap N$, which is equal to $\iota^{-1}(S)$, is a recognisable subset of N by Proposition 2.11 and, for the same reason, $\pi^{-1}(S \cap N)$ is a recognisable subset of A^*. By Kleene's theorem, the language $K' = K \cap \pi^{-1}(S \cap N)$ is also rational. It follows by Proposition 1.1 that $\iota(\pi(K'))$ is a rational subset of M. Now, since π is surjective, Corollary I.1.11 and Proposition I.1.19 give the relations

$$\pi(K') = \pi(K \cap \pi^{-1}(S \cap N)) = \pi(K) \cap (S \cap N)$$

and since ι is injective, one has by Proposition I.1.17:

$$\iota(\pi(K')) = \iota(\pi(K) \cap (S \cap N)) = \iota(\pi(K)) \cap \iota(S \cap N) = R \cap S$$

It follows that $R \cap S$ is a rational subset of M. □

The next theorem gives a description of the recognisable subsets of a finite product of monoids. Note that this result does not extend to finite products of semigroups.

[1] Apologies for the forward reference, but it makes the demonstration more fluent.

Theorem 2.16 (Mezei). *Let $M_1, \ldots, M_n$ be monoids and let $M = M_1 \times \cdots \times M_n$. A subset of M is recognisable if and only if it is a finite union of subsets of the form $R_1 \times \cdots \times R_n$, where each R_i is a recognisable subset of M_i.*

Proof. Since R_i is recognisable, there exists a morphism α_i from M_i onto some finite monoid F_i such that $\alpha_i^{-1}(\alpha(R_i)) = R_i$. Let $\alpha : M \to F_1 \times \cdots \times F_n$ be the product morphism defined by $\alpha(m_1, \ldots, m_n) = (\alpha_1(m_1), \ldots, \alpha_n(m_n))$. The relations

$$\begin{aligned}\alpha^{-1}(\alpha(R_1 \times \cdots \times R_n)) &= \alpha_1^{-1}(\alpha_1(R_1)) \times \cdots \times \alpha_n^{-1}(\alpha_n(R_n)) \\ &= R_1 \times \cdots \times R_n\end{aligned}$$

show that $R_1 \times \cdots \times R_n$ is recognisable. Furthermore, Corollary 2.13 shows that any finite union of subsets of this form is recognisable.

Consider now a recognisable subset R of M. Then there exists a morphism α from M onto a finite monoid F such that $R = \alpha^{-1}(\alpha(R))$. For $1 \leqslant i \leqslant n$, let us define a morphism $\beta_i : M_i \to F$ by setting $\beta_i(m_i) = \alpha(1, \ldots, 1, m_i, 1, \ldots, 1)$. We also define a morphism $\beta : M \to F^n$ by setting

$$\beta(m_1, \ldots, m_n) = (\beta_1(m_1), \ldots, \beta_n(m_n)).$$

Setting

$$Q = \{(x_1, \ldots, x_n) \in F^n \mid x_1 \cdots x_n \in \alpha(R)\}$$

and observing that $\beta_1(m_1) \cdots \beta_n(m_n) = \alpha(m_1, \ldots, m_n)$, one gets

$$\begin{aligned}\beta^{-1}(Q) &= \{(m_1, \ldots, m_n) \in M \mid \beta_1(m_1) \cdots \beta_n(m_n) \in \alpha(R)\} \\ &= \{(m_1, \ldots, m_n) \in M \mid \alpha(m_1, \ldots, m_n) \in \alpha(R)\} \\ &= \alpha^{-1}\alpha(R) = R\end{aligned}$$

Furthermore, $\beta^{-1}(x_1, \ldots, x_n) = \beta_1^{-1}(x_1) \times \cdots \times \beta_n^{-1}(x_n)$ and therefore

$$R = \beta^{-1}(Q) = \bigcup_{(x_1, \ldots, x_n) \in Q} \beta_1^{-1}(x_1) \times \cdots \times \beta_n^{-1}(x_n)$$

Since the sets $\beta_i^{-1}(x_i)$ are by construction recognisable subsets of M_i, the desired decomposition is obtained. □

One of the most important applications of Theorem 2.16 is the fact that the product of two recognisable relations over finitely generated free monoids is recognisable. Let $A_1, \ldots, A_n$ be finite alphabets. Then the monoid $A_1^* \times \cdots \times A_n^*$ is finitely generated, since it is generated by the finite set

$$\{(1, \ldots, 1, a_i, 1, \ldots, 1) \mid a_i \in A_i, 1 \leqslant i \leqslant n\}.$$

Proposition 2.17. *Let $A_1, \ldots, A_n$ be finite alphabets. The product of two recognisable subsets of $A_1^* \times \cdots \times A_n^*$ is recognisable.*

Proof. Let $M = A_1^* \times \cdots \times A_n^*$ and let X and Y be two recognisable subsets of M. By Theorem 2.16, X is a finite union of subsets of the form $R_1 \times \cdots \times R_n$, where R_i is a recognisable subset of A_i^* and Y is a finite union of subsets of the form $S_1 \times \cdots \times S_n$, where S_i is a recognisable subset of A_i^*. It follows that XY is a finite union of subsets of the form $R_1S_1 \times \cdots \times R_nS_n$. Since A_i^* is a free monoid, the sets R_i and S_i are rational by Kleene's theorem. It follows that R_iS_i is rational and hence recognisable. Finally, Theorem 2.16 shows that XY is recognisable. □

3 Connection with automata

The case of the free monoid is of course the most important. There is a natural way to associate a finite monoid with a finite automaton. We start by explaining this construction for deterministic automata, which is a bit easier than the general case.

3.1 Transition monoid of a deterministic automaton

Let $\mathcal{A} = (Q, A, E, q_-, F)$ be a deterministic automaton. Each letter $a \in A$ defines a partial transformation $q \to q \cdot a$ on Q, which maps each state q onto the unique state $q \cdot a$ (when it exists) such that $(q, a, q \cdot a) \in E$. If there is no such state, $q \cdot a$ is undefined.

More generally, each word u defines a partial transformation $q \to q \cdot u$ on Q, which maps each state q onto the end state of the unique path with label u and origin q (when it exists). One can also define $q \cdot u$ by induction on the length of u. First, the empty word defines the identity transformation: for each state q, $q \cdot 1 = q$. Next, for every word $u \in A^+$ and for every letter $a \in A$, one sets $q \cdot (ua) = (q \cdot u) \cdot a$ if $(q \cdot u)$ and $(q \cdot u) \cdot a$ are defined.

Example 3.1. Let $\mathcal{A}$ be the automaton of Figure III.3.4. One has $1 \cdot ba = 2$, $2 \cdot ba = 1$ and $3 \cdot ba = 5$ and for $q = 4, 5$, the value of $q \cdot ba$ is undefined. Similarly, the partial transformation associated to $babbb$ is defined on its domain $\{1, 2\}$ by $1 \cdot babbb = 2$ and $2 \cdot babbb = 1$.

The partial function $(q, a) \to q \cdot a$ from $Q \times A$ to Q is called the *transition function* of $\mathcal{A}$. If $\mathcal{A}$ is complete, it is a total function. Note that the set E of transitions completely determines the transition function and can also be recovered from it. For this reason, a deterministic automaton is often given as a triple $(Q, A, \cdot, q_-, F)$, where "$\cdot$" denotes the transition function.

It follows from the definitions that the function which maps a word u onto the partial transformation $q \to q \cdot u$ defines a morphism from A^* to the monoid $\mathcal{F}(Q)$ of partial transformations on Q. The range of this map is a monoid, called the *transition monoid* of $\mathcal{A}$.

We now present an algorithm to compute a presentation of the transition monoid of a finite automaton. Rather than giving a formal description of the algorithm, we shall explain it step by step on an example. Consider the finite deterministic automaton represented in Figure 3.1.

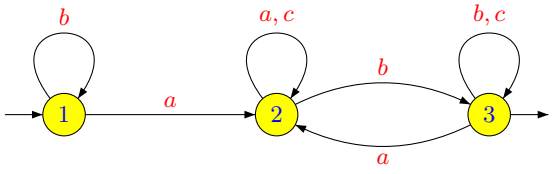


Figure 3.1. A deterministic automaton.

We start our computation by giving the generators of its transition monoid.

	1	2	3
a	2	2	2
b	1	3	3
c	-	2	3

Thus a maps all states onto state 2 and c defines a partial identity: it maps states 2 and 3 onto themselves and is undefined on state 1.

The transition monoid is equal to the transformation monoid generated by these generators. In order to compute it, we proceed as follows. We first fix a total order on the alphabet, for instance $a < b < c$, but any total order would do. We also let $<$ denote the shortlex order induced by this total order (see Section III.1.2 for the definition). We maintain simultaneously a list of elements, in increasing order for $<$, and a list of rewriting rules of the form $u \to v$, where $v < u$.

We compute the elements of the transition monoid by induction on the rank of the words in the shortlex order. The partial transformation associated with the empty word 1 is the identity. To remember this, we add a 1 in the first column of the first row.

1	1	2	3

The partial transformations associated with the letters are the three generators and thus we get the following table for the words of length $\leqslant 1$:

1	1	2	3
a	2	2	2
b	1	3	3
c	-	2	3

The general principle is now the following. Each time we consider a new word w, we first try to use an existing rewriting rule to reduce w. If this is possible, we simply consider the next word in the shortlex order. If this is not possible, we compute the partial transformation associated with w and check whether or not this partial transformation already appeared in our list of partial transformations. If it already appeared as u, we add $w \to u$ to the list of the rewriting rules, otherwise, we add w to the list of elements. Let us also mention a convenient trick to compute the partial transformation associated with w when w is irreducible. Suppose that $w = px$, where x is a letter. Since w is irreducible, so is p and hence p is certainly in the list of elements. Therefore we already know the partial transformation associated with p and it suffices to compose it with x to get the partial transformation associated with w.

Let us go back to our example. Since we already considered the words of length $\leqslant 1$, we now consider the first word of length 2, namely aa. The partial transformation associated with aa can be computed directly by following the transition on Figure 3.1, or by using the table below. One can see that aa defines the same partial transformation as a and thus we add the rewriting rule $aa \to a$ to our list. Next, we compute ab. This transformation maps all states to 3 and is a new partial transformation. We then proceed by computing ac, which defines the same partial transformation as a, and then, ba, bb (which give the new rewriting rules $ba \to a$ and $bb \to b$), bc and ca (which define new partial

transformations) and cb and cc (which give the new rewriting rules $cb \to bc$ and $cc \to c$). At this stage our two lists look like this:

1	1	2	3
a	2	2	2
b	1	3	3
c	-	2	3
ab	3	3	3
bc	-	3	3
ca	-	2	2

Rewriting rules

$aa \to a$	$bb \to b$
$ac \to a$	$cb \to bc$
$ba \to a$	$cc \to c$

The next step is to compute the partial transformations corresponding to the words of length 3, but it suffices to consider the words starting with ab, bc or ca, since the other words can be reduced using the rewriting rules we have already produced. Now aba can be reduced by using $ba \to a$, abb can be reduced by using $bb \to b$. The word abc defines the same partial transformation as ab, which gives the new rule $abc \to ab$. Similarly, bca defines the same partial transformation as ca, giving the rule $bca \to ca$. Moreover, bcb, bcc and caa can be reduced using $cb \to bc$, $cc \to c$ and $aa \to a$ respectively. Finally, cab defines the same partial transformation as bc, which gives the new rule $abc \to bc$ and cac can be reduced by using the rule $ac \to a$.

No word of length 3 defines a new partial transformation and thus the algorithm terminates. A presentation of the transition monoid of $\mathcal{A}$ is given in the following table:

1	1	2	3
a	2	2	2
b	1	3	3
c	-	2	3
ab	3	3	3
bc	-	3	3
ca	-	2	2

Rewriting rules

$aa \to a$	$bb \to b$	$abc \to ab$
$ac \to a$	$cb \to bc$	$bca \to ca$
$ba \to a$	$cc \to c$	$cab \to bc$

One can show that the rewriting system obtained at the end of the algorithm is converging, which means that every word has a unique reduced form. In particular, the order in which the reduction rules are applied does not matter. The rewriting system can be used to compute the product of two elements. For instance $abbc \to abc \to ab$ and thus $(ab)(bc) = ab$.

A presentation of the transition monoid is obtained by replacing the symbol $\to$ by $=$. In our example, the presentation would be $\langle \{a, b, c\} \mid aa = a, bb = b, abc = ab, ac = a, cb = bc, bca = ca, ba = a, cc = c, cab = bc \rangle$.

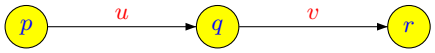
3.2 Transition monoid of a nondeterministic automaton

Let now $\mathcal{A} = (Q, A, E, I, F)$ be a nondeterministic finite automaton. To each word $u \in A^*$, there corresponds a relation on Q, denoted by $\mu(u)$, and defined by $(p, q) \in \mu(u)$ if there exists a path from p to q with label u.

Proposition 3.18. *The function μ is a morphism from A^* to the monoid of relations on Q.*

Proof. Clearly, $\mu(1)$ is the identity relation. Let now u and v be two words of A^*. Then one has $(p,r) \in \mu(u)\mu(v)$ if and only if there is a state q such that $(p,q) \in \mu(u)$ and $(q,r) \in \mu(v)$. By definition of μ, this means there is a path from p to q with label u and path from q to r with label v. Therefore, there is a path from p to r with label uv and $(p,r) \in \mu(uv)$.

Conversely, suppose that $(p,r) \in \mu(uv)$. Then there is a path from p to r with label uv. Let q the state reached on this path after reading u:

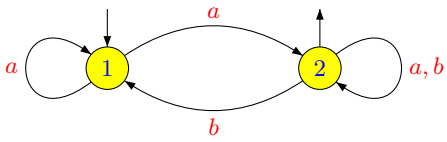


Then $(p,q) \in \mu(u)$ and $(q,r) \in \mu(v)$ and hence $(p,r) \in \mu(u)\mu(v)$. □

The monoid $\mu(A^*)$ is called the *transition monoid* of $\mathcal{A}$ and denoted by $M(\mathcal{A})$. For practical computation, it can be conveniently represented as a monoid of Boolean matrices of order $|Q|\times|Q|$. In this case, $\mu(u)$ can be identified with the matrix defined by

$$\mu(u)_{p,q} = \begin{cases} 1 & \text{if there exists a path from } p \text{ to } q \text{ with label } u \\ 0 & \text{otherwise} \end{cases}$$

Example 3.2. If $\mathcal{A}$ is the automaton below, one gets



$$\mu(a) = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \qquad \mu(b) = \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix} \qquad \mu(aa) = \mu(a)$$
$$\mu(ab) = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \qquad \mu(ba) = \mu(bb) = \mu(b)$$

Thus the transition monoid of $\mathcal{A}$ is the monoid of Boolean matrices

$$\mu(A^*) = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \right\}.$$

3.3 Monoids versus automata

We now show that our definition of recognisable sets is equivalent with the standard definition using automata.

Proposition 3.19. *If a finite automaton recognises a language L, then its transition monoid recognises L.*

Proof. Let $\mathcal{A}$ be a finite automaton recognising L and let $\mu : A^* \to M(\mathcal{A})$ be the natural morphism from A^* onto the transition monoid of $\mathcal{A}$. Observe now

that a word u is recognised by $\mathcal{A}$ if and only if $(p, q) \in \mu(u)$ for some initial state p and some final state q. It follows that $L = \mu^{-1}(P)$ where

$$P = \{m \in M(\mathcal{A}) \mid m_{p,q} = 1 \text{ for some initial state } p \text{ and some final state } q\}$$

It follows that μ recognises L. □

Example 3.3. Let $A = \{a, b\}$ and let $\mathcal{A}$ be the (incomplete) deterministic automaton represented below.

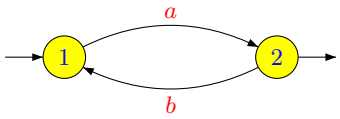


Figure 3.2. A deterministic automaton.

It is easy to see that $\mathcal{A}$ recognises the language $(ab)^*a$. The transition monoid M of $\mathcal{A}$ contains six elements which correspond to the words 1, a, b, ab, ba and aa. Furthermore aa is a zero of S and thus can be denoted by 0. The other relations defining M are $aba = a$, $bab = b$ and $bb = 0$.

1	a	b	aa	ab	ba
1	2	–	–	1	–
2	–	1	–	–	2

One recognises the monoid B_2^1 defined on page 18.

Converting a monoid to an automaton is also very easy.

Proposition 3.20. *If a language is recognised by a finite monoid, then it is also recognised by a finite deterministic automaton.*

Proof. Let L be a language recognised by a finite monoid M. Then there is a morphism $\varphi : A^* \to M$ and a subset P of M such that $L = \varphi^{-1}(P)$. Define a deterministic automaton $\mathcal{A} = (M, A, \cdot, 1, P)$, where, for every $s \in M$ and $a \in A$, $s \cdot a = s\varphi(a)$. Now, a word u is accepted by $\mathcal{A}$ if and only if $1 \cdot u \in P$. Since $1 \cdot u = \varphi(u)$, this condition means $\varphi(u) \in P$ or $u \in \varphi^{-1}(P)$. Since $L = \varphi^{-1}(P)$, we conclude that $\mathcal{A}$ recognises L. □

Example 3.4. Let $\varphi : \{a, b\}^* \to B_2^1 = \{1, a, b, ab, ba, 0\}$ be the morphism defined by $\varphi(a) = a$ and $\varphi(b) = b$ and let $P = \{1, ab\}$. By applying the algorithm described above, one gets the automaton pictured in Figure 3.3, which recognises $(ab)^*$.

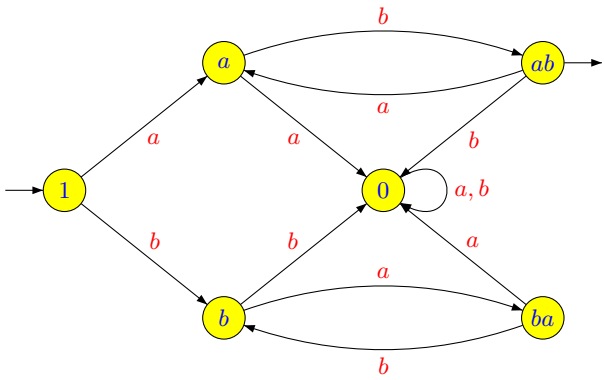


Figure 3.3. The automaton associated with φ.

We can now summarise the results of this section.

Theorem 3.21. *Let L be a language. The following conditions are equivalent*

(1) *L is recognised by a finite deterministic automaton,*

(2) *L is recognised by a finite automaton,*

(3) *L is recognised by a finite monoid.*

Proof. (1) implies (2) is clear and (2) implies (3) follows from Proposition 3.19. Finally, (3) implies (1) follows from Proposition 3.20. □

This result implies that any finite automaton is equivalent to a deterministic one, a good illustration of the power of the algebraic approach. Indeed, the usual proof requires the subset construction, but the proof is straightforward with monoids.

4 The syntactic monoid

The syntactic congruence is one of the key notions of this chapter. Roughly speaking, it is the monoid analog of the notion of a minimal automaton, but it can be defined for any subset of a monoid.

4.1 Definitions

The main definition was already given in Section II.3.7. Given a subset L of M the *syntactic congruence* of L in M is the relation $\sim_L$ defined on M by $u \sim_L v$ if and only if, for all $x, y \in M$,

$$xuy \in L \Longleftrightarrow xvy \in L \tag{4.1}$$

The quotient $M/\sim_L$ is the *syntactic monoid* of L and the natural morphism $\eta_L : M \to M/\sim_L$ is the *syntactic morphism* of L. Finally, the set $\eta_L(L)$ is called the *syntactic image* of L. The syntactic monoid of a language L is often denoted by $M(L)$ and we shall freely use this notation.

The syntactic congruence is characterised by the following property.

Proposition 4.22. *The syntactic congruence of L is the coarsest congruence that saturates L. Furthermore, a congruence $\sim$ saturates L if and only if $\sim_L$ is coarser than $\sim$.*

Proof. We first need to show that $\sim_L$ saturates L. Suppose that $u \in L$ and $u \sim_L v$. Then, one has, by taking $x = y = 1$ in (4.1),

$$u \in L \Longleftrightarrow v \in L$$

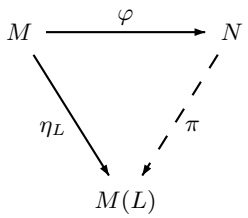
and thus $v \in L$. It follows that η_L recognises L.

Suppose now that a congruence $\sim$ saturates L and that $u \sim v$. Since $\sim$ is a congruence, we also have $xuy \sim xvy$ for every $x, y \in M$. Since $\sim$ saturates L, it follows that $xuy \in L$ if and only if $xvy \in L$ and hence $u \sim_L v$. Therefore, $\sim_L$ is coarser than $\sim$.

Finally, if $\sim_L$ is coarser than $\sim$, the condition $u \sim v$ implies $u \sim_L v$. Therefore, if $u \in L$, then $v \in L$ and $\sim$ saturates L. □

It is sometimes convenient to state this result in terms of morphisms:

Corollary 4.23. *Let $\varphi : M \to N$ be a surjective morphism of monoids, let L be a subset of M and let $\eta_L : M \to M(L)$ be the syntactic morphism of L. Then φ recognises L if and only if there is a surjective morphism $\pi : N \to M(L)$ such that $\eta_L = \pi \circ \varphi$.*



One says that η_L factors through φ.

Corollary 4.23 is often used in the following simpler form, which implies that the syntactic monoid of a subset is the smallest monoid recognising this subset.

Corollary 4.24. *Let L be a subset of a monoid M. A monoid N fully recognises L if and only if the syntactic monoid of L is a quotient of N.*

Let us state another useful result:

Proposition 4.25. *Let $\varphi : M \to N$ be a surjective morphism and let L be a subset of N. Then the syntactic monoid of L in N is equal to the syntactic monoid of $\varphi^{-1}(L)$ in M.*

Proof. Let $u, v \in M$. We claim that $u \sim_{\varphi^{-1}(L)} v$ if and only if $\varphi(u) \sim_L \varphi(v)$. Indeed, the condition $u \sim_{\varphi^{-1}(L)} v$ means that, for all $x, y \in M$,

$$xuy \in \varphi^{-1}(L) \Longleftrightarrow xvy \in \varphi^{-1}(L)$$

or, equivalently

$$\varphi(xuy) \in L \Longleftrightarrow \varphi(xvy) \in L$$

Since φ is surjective, this latter condition exactly means that $\varphi(u) \sim_L \varphi(v)$, proving the claim. It follows that $M/\sim_{\varphi^{-1}(L)}$ is isomorphic to $N/\sim_L$, which gives the result. □

4.2 The syntactic monoid of a language

Corollary 4.23 applies in particular when M is a free monoid A^*. In this case, one can characterise the syntactic monoid in terms of division.

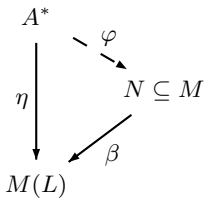
Proposition 4.26. *Let L be a language of A^* and let $M(L)$ be the syntactic monoid of L.*

(1) *A monoid M recognises L if and only if $M(L)$ divides M.*

(2) *If M recognises L and if M divides N, then N recognises L.*

Proof. (1) Let $\varphi : A^* \to M$ be a morphism recognising L. Then the morphism $\varphi : A^* \to \varphi(A^*)$ also recognises L and by Corollary 4.23, $M(L)$ is a quotient of $\varphi(A^*)$ and thus divides M.

Let $\eta : A^* \to M(L)$ be the syntactic morphism of L. If $M(L)$ divides M, there is a submonoid N of M and a surjective morphism $\beta : N \to M(L)$. According to Corollary II.5.30, there exists a morphism $\varphi : A^* \to M$ such that $\eta = \beta \circ \varphi$. Let $P = \beta^{-1}(\eta(L))$.



Then

$$L = \eta^{-1}\eta(L) = \varphi^{-1}(\beta^{-1}(\eta(L)) = \varphi^{-1}(P)$$

and hence M recognises L.

(2) If M recognises L, then by (1), $M(L)$ divides M. Thus if M divides N, $M(L)$ also divides N and by (1) again, N recognises L. □

It is interesting to translate the results of Section 2.2 in terms of syntactic monoids. In the next proposition, we let $M(L)$ denote the syntactic monoid of a language L.

Proposition 4.27. *Let L, L_1, L_2 and K be languages of A^*. Then the following properties hold:*

(1) $M(L^c) = M(L)$,

(2) $M(L_1 \cap L_2)$ *divides* $M(L_1) \times M(L_2)$,

(3) $M(L_1 + L_2)$ *divides* $M(L_1) \times M(L_2)$,

(4) $M(LK^{-1})$ *and* $M(K^{-1}L)$ *divide* $M(L)$,

(5) *If $\varphi : B^* \to A^*$ is a morphism, then $M(\varphi^{-1}(L))$ divides $M(L)$.*

Proof. This is an immediate consequence of Propositions 2.9, 2.10, 2.11, 2.12 and 4.22. □

Note that the proof of these results is notably shorter than their automata theoretic counterpart.

4.3 Computation of the syntactic monoid of a language

The easiest way to compute the syntactic monoid of a recognisable language L is to first compute its minimal (deterministic) automaton and then to apply the following result.

Proposition 4.28. *The syntactic monoid of a recognisable language is equal to the transition monoid of its minimal automaton.*

Proof. Let $\mathcal{A} = (Q, A, \cdot, q_-, F)$ be the minimal automaton of a recognisable language L of A^* and let M be its transition monoid. It suffices to verify that two words $u, v \in A^*$ satisfy $u \sim_L v$ if and only if they define the same transition in M.

Suppose that $u \sim_L v$ and let $p \in Q$. Since $\mathcal{A}$ is trimmed, there exists a word $x \in A^*$ such that $q_- \cdot x = p$. Now since $u \sim_L v$, one has, for all $y \in A^*$,

$$xuy \in L \Longleftrightarrow xvy \in L$$

that is,

$$q_- \cdot xuy \in F \Longleftrightarrow q_- \cdot xvy \in F$$

or yet

$$(p \cdot u) \cdot y \in F \Longleftrightarrow (p \cdot v) \cdot y \in F.$$

Therefore, the states $p \cdot u$ and $p \cdot v$ are Nerode equivalent and hence equal by Corollary III.4.20. Thus u and v define the same transition in M.

Suppose now that u and v define the same transition in M. Then if $xuy \in L$, then $q_- \cdot xuy \in F$. Since xuy and xvy define the same transition in M, one also gets $q_- \cdot xvy \in F$, and hence $xvy \in L$. □

5 Recognition by ordered structures

We have seen that a subset of a monoid and its complement have the same syntactic monoid. Similarly, a language and its complement have the same complete minimal automaton. Although this property is usually very convenient, there are cases where it is desirable to distinguish between a subset and its complement. Introducing a partial order on monoids solves this problem in an elegant way.

5.1 Ordered automata

An *ordered automaton* is a deterministic automaton $\mathcal{A} = (Q, A, \cdot, q_-, F)$ equipped with a partial order $\leqslant$ on Q such that, for all $p, q \in Q$ and $a \in A$, $p \leqslant q$ implies $p \cdot a \leqslant q \cdot a$. This implies in particular that, for all $u \in A^*$, $p \cdot u \leqslant q \cdot u$.

There is a natural way to define a preorder on the set of states of any deterministic automaton. Let $\mathcal{A} = (Q, A, \cdot, q_-, F)$ be a deterministic automaton. Define a relation $\leqslant$ on Q by setting $p \leqslant q$ if and only if, for all $u \in A^*$, $p \cdot u \in F$ implies $q \cdot u \in F$. This preorder turns out to be an order if $\mathcal{A}$ is minimal and complete.

Proposition 5.29. *If $\mathcal{A}$ is a minimal complete deterministic automaton, the relation $\leqslant$ is a partial order on Q.*

Proof. The relation $\leqslant$ is clearly reflexive and transitive. Suppose that $p \leqslant q$ and $q \leqslant p$. Then, for all $u \in A^*$, $p \cdot u \in F \Leftrightarrow q \cdot u \in F$. Since $\mathcal{A}$ is minimal, this implies $p = q$. Thus $\leqslant$ is an order.

Furthermore, if $p \leqslant q$, then for all $a \in A$, $p \cdot a \leqslant q \cdot a$ since, for all $u \in A^*$, $p \cdot au \in F$ implies $q \cdot au \in F$. □

Example 5.1. Consider the minimal complete automaton of $(ab)^*$, represented in Figure III.4.14. The order on the set of states is $0 < 1$ and $0 < 2$. Indeed, one has $0 \cdot u = 0$ for all $u \in A^*$ and thus, the formal implication

$$0 \cdot u \in F \Rightarrow q \cdot u \in F$$

holds for any $q \in Q$. One can verify that there is no other relations among the states of Q. For instance, 1 and 2 are incomparable since $1 \cdot ab = 1 \in F$ but $2 \cdot ab = 0 \notin F$ and $1 \cdot b = 0 \notin F$ but $2 \cdot b = 1 \in F$.

Example 5.2. Consider the minimal complete automaton of aA^*b, represented in Figure III.4.15. The order on the set of states is $0 < 1 < 2 < 3$.

5.2 Recognition by ordered monoids

A *congruence on an ordered monoid* $(M, \leqslant)$ is a stable preorder which is coarser than $\leqslant$. In particular, the order relation $\leqslant$ is itself a congruence. If $\preccurlyeq$ is a congruence on $(M, \leqslant)$, then the equivalence relation $\sim$ associated with $\preccurlyeq$ is a congruence on M. Furthermore, there is a well-defined stable order on the quotient set $M/\!\sim$, given by

$$[s] \leqslant [t] \quad \text{if and only if } s \preccurlyeq t$$

Thus $(M/\!\sim, \leqslant)$ is an ordered monoid, also denoted by $M/\!\preccurlyeq$.

Let $\varphi : M \to N$ be a surjective morphism of ordered monoids. A subset Q of M is *recognised* by φ if there exists an upper set P of N such that

$$Q = \varphi^{-1}(P)$$

This condition implies that Q is an upper set of M and that

$$\varphi(Q) = \varphi(\varphi^{-1}(P)) = P$$

By extension, a subset Q of M is said to be *recognised* by an ordered monoid N if there exists a surjective morphism of ordered monoids from M onto N that recognises Q.

It is sometimes convenient to formulate this definition in terms of congruences. Let M be an ordered monoid and let $\preccurlyeq$ a congruence on M. A subset Q of M is said to be *recognised* by $\preccurlyeq$ if, for every $q \in Q$, $q \preccurlyeq p$ implies $p \in Q$. It is easy to see that a surjective morphism of ordered monoids φ recognises Q if and only if the nuclear congruence $\preccurlyeq_\varphi$ recognises Q.

How does this definition relate to the standard definition using monoids? Simply by considering a monoid as an ordered monoid, with the equality relation as an order relation.

Example 5.3. Consider the monoid $B_2^1 = \{1, a, b, ab, ba, 0\}$ and let $\pi : \{a, b\}^* \to M$ be the morphism defined by $\pi(a) = a$ and $\pi(b) = b$. Let us define an order on B_2^1 by setting $ab \leqslant 1$, $ba \leqslant 1$, and $0 \leqslant x$ for all $x \in B_2^1$. Then the set $I = \{1, ab\}$ is an upper set and $\pi^{-1}(I) = (ab)^*$. Thus the language $(ab)^*$ is recognised by the ordered monoid $(B_2^1, \leqslant)$.

It suffices to reverse the order to recognise the complement of a language.

Proposition 5.30. *Let L be a language of A^*. If L is recognised by an ordered monoid $(M, \leqslant)$, then L^c is recognised by the ordered monoid $(M, \geqslant)$.*

Proof. Let $\varphi : A^* \to M$ be a morphism and let P be an upper set of $(M, \leqslant)$ such that $\varphi^{-1}(P) = L$. Then $\varphi^{-1}(P^c) = L^c$ and it suffices to show that P^c is a $\geqslant$-upper set. Suppose that $x \in P^c$ and $x \geqslant y$. If $y \in P$, then $x \in P$, since P is a $\leqslant$-upper set. Thus $y \in P^c$, which concludes the proof. □

5.3 Syntactic order

First note that, if $(M, \leqslant)$ is an ordered monoid, the congruence $\leqslant$ recognises every upper set of M. The syntactic congruence of an upper set Q of M is the coarsest congruence among the congruences on M that recognise Q.

Let M be an ordered monoid and let P be an upper set of M. Define a relation $\preccurlyeq_P$ on M by setting $u \preccurlyeq_P v$ if and only if, for every $x, y \in M$,

$$xuy \in P \Rightarrow xvy \in P$$

One can show that the relation $\preccurlyeq_P$ is a congruence of ordered monoids on M that recognises P. This congruence is called the *syntactic congruence* of P in M.

The ordered monoid $(M(P), \leqslant) = M/\preccurlyeq_P$ is the *syntactic ordered monoid* of P, the order relation on $M(P)$ the *syntactic order* of P and the quotient morphism η_P from M onto $M(P)$ the *syntactic morphism* of P.

The syntactic congruence is characterised by the following property.

Proposition 5.31. *The syntactic congruence of P is the coarsest congruence that recognises P. Furthermore, a congruence $\preccurlyeq$ recognises P if and only if $\preccurlyeq_P$ is coarser than $\preccurlyeq$.*

It is sometimes convenient to state this result in terms of morphisms:

Corollary 5.32. *Let $\varphi : M \to N$ be a surjective morphism of ordered monoids and let P be an upper set of M. The following properties hold:*

(1) *The morphism φ recognises P if and only if η_P factors through it.*

(2) *Let $\pi : N \to R$ be a surjective morphism of ordered monoids. If $\pi \circ \varphi$ recognises P, then φ recognises P.*

5.4 Computation of the syntactic ordered monoid

We already know that the syntactic monoid of a language is equal to the transition monoid of its minimal automaton. Its syntactic order can be computed in two different ways: either directly on the syntactic monoid or by using the order on the complete minimal automaton.

Let L be a language, let $\eta : A^* \to M$ be its syntactic monoid and let $P = \eta(L)$. We define a preorder relation $\leqslant_L$ on A^* by setting $u \leqslant_L v$ if and only if, for every $s, t \in A^*$,

$$sut \in L \Rightarrow svt \in L$$

Recall that the syntactic order $\leqslant_P$ is the partial order on M defined as follows : $u \leqslant_P v$ if and only if, for every $s, t \in M$,

$$sut \in P \Rightarrow svt \in P$$

Let $\mathcal{A} = (Q, A, \cdot, q_-, F)$ be the minimal automaton of L and let $\leqslant$ be the natural order on Q. The next proposition gives the relations between $\leqslant_L$, $\leqslant_P$ and $\leqslant$.

Proposition 5.33. *Let u and v be two words of A^*. The following conditions are equivalent:*

(1) $u \leqslant_L v$,
(2) $\eta(u) \leqslant_P \eta(v)$,
(3) *for all $q \in Q$, $q \cdot u \leqslant q \cdot v$.*

Proof. (1) implies (2). Suppose that $u \leqslant_L v$ and let $s, t \in M$ be such that $s\eta(u)t \in P$. Since η is surjective, there exist some words x, y such that $\eta(x) = s$ and $\eta(y) = t$. Observing that $s\eta(u)t = \eta(x)\eta(u)\eta(y) = \eta(xuy)$, one gets $\eta(xuy) \in P$ and hence $xuy \in L$ as η recognises L. Condition (1) implies that $xvy \in L$, or equivalently, $\eta(xvy) \in P$. By a similar argument, we get $\eta(xvy) = s\eta(v)t$ and thus $s\eta(v)t \in P$, which proves (2).

(2) implies (1). Suppose that $\eta(u) \leqslant_P \eta(v)$ and let $x, y \in A^*$ be such that $xuv \in L$. Then $\eta(xuv) \in P$ and since $\eta(xuv) = \eta(x)\eta(u)\eta(y)$, one gets by (2) $\eta(x)\eta(v)\eta(y) \in P$. Observing again that $\eta(x)\eta(v)\eta(y) = \eta(xvy)$, one concludes that $\eta(xvy) \in P$ and finally $xvy \in L$. Thus $u \leqslant_L v$.

(1) implies (3). Suppose that $u \leqslant_L v$ and let q be a state of Q. Since $\mathcal{A}$ is trimmed, there is a word x such that $q_- \cdot x = q$. Let y be a word such that $(q \cdot u) \cdot y \in F$. Then $q_- \cdot xuy \in F$ and thus $xuy \in L$. Since $u \leqslant_L v$, one has $xvy \in L$ and hence $(q \cdot v) \cdot y \in F$. Thus $q \cdot u \leqslant q \cdot v$.

(3) implies (1). Assume that, for every $q \in Q$, $q \cdot u \leqslant q \cdot v$. Let $x, y \in A^*$. If $xuy \in L$, then $q_- \cdot xuy \in F$ and since $q_- \cdot xu \leqslant q_- \cdot xv$, one gets $q_- \cdot xvy \in F$ and finally $xvy \in L$. Thus $u \leqslant_L v$. □

Corollary 5.34. *The syntactic ordered monoid of a recognisable language is equal to the transition monoid of its ordered minimal automaton.*

Example 5.4. Let $\mathcal{A}$ be the deterministic automaton of $(ab)^*$ (Example 5.1). Its transition monoid was calculated in Example 3.3. Its syntactic monoid and its syntactic order are represented below (an arrow from u to v means that $u < v$).

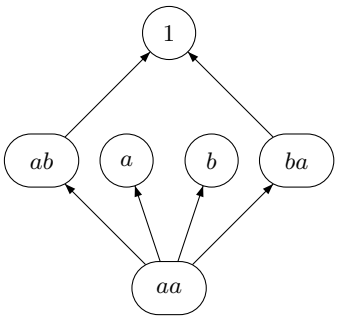
Elements

	1	2
1	1	2
a	2	0
b	0	1
aa	0	0
ab	1	0
ba	0	2

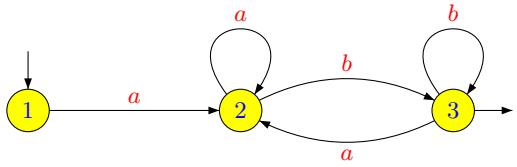
Relations

$bb = aa$
$aba = a$
$bab = b$

Syntactic order



Example 5.5. The ordered minimal automaton of the language aA^*b was computed in Example 5.2. This automaton is represented again below, but the sink state 0 is omitted. The order on the states is $0 < 1 < 2 < 3$. The elements of the syntactic monoid, the relations defining it and the syntactic order are also presented in the tables below.



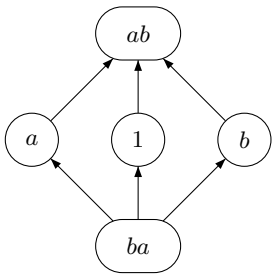
Elements

	1	2	3
1	1	2	3
a	2	2	2
b	0	3	3
ab	3	3	3
ba	0	2	2

Relations

$aa = a$
$bb = b$
$aba = a$
$bab = b$

Syntactic order



6 Exercises

Section 1

Exercise 1. (Difficult). Show that the rational subsets of a finitely generated commutative monoid form a Boolean algebra.

Exercise 2. (Difficult). Show that the rational subsets of a finitely generated free group form a Boolean algebra.

Section 2

Exercise 3. Show that the set $\{0\}$ is a rational subset of the additive group $\mathbb{Z}$ but is not recognisable.

Exercise 4. Describe the rational [recognisable] subsets of the additive monoids $\mathbb{N}$ and $\mathbb{Z}$.

Exercise 5. (Difficult). Let G be a group and let H be a subgroup of G.

(1) Show that H is a recognisable subset of G if and only if H has finite index in G.
(2) Show that H is a rational subset of G if and only if H is finitely generated.

Section 3

Exercise 6. Let $\mathcal{A}$ be a complete n-state deterministic automaton. Show that the transition monoid of $\mathcal{A}$ has at most n^n elements. Can this bound be reached?

Suppose now that $\mathcal{A}$ is an n-state nondeterministic automaton. Show that the transition monoid of $\mathcal{A}$ has at most 2^{n^2} elements. Can this bound be reached?

Exercise 7. A *group language* is a language whose syntactic monoid is a finite group. Show that a language is a group language if and only if it is recognised by a permutation automaton (i.e. for each letter a, the map $q \to q \cdot a$ is a permutation of the set of states).

Section 4

Exercise 8. Compute the syntactic monoid of the following languages on the alphabet $\{a, b\}^*$:

(1) $\{u \mid |u| \text{ is odd}\}$
(2) a^*
(3) $\{aba, b\}$
(4) $(ab)^*$
(5) $(aba)^*$
(6) $A^*aA^*aA^*bA^*$
(7) A^*aaA^*
(8) A^*abA^*
(9) A^*abaA^*
(10) $(a(ab)^*b)^*$
(11) $(aa + bb)^*$
(12) $(ab + ba)^*$
(13) $(a + bab)^*$
(14) $(a + bb + aba)^*$

Exercise 9. Let L be a language of A^* and let $u, v, z \in A^*$. Show that if $z \sim_L zu \sim_L zv$, then $zu^2 \sim_{LaA^*} zu$ and $zuv \sim_{LaA^*} zvu$.

Exercise 10. Show that if L is a group language, then the idempotents of the syntactic monoid of LaA^* commute.

Exercise 11. Suppose that the syntactic monoid of a language L is the full transformation semigroup $\mathcal{T}_n$. Let $\tilde{L} = \{\tilde{u} \mid u \in L$. Show that the minimal automaton of $\tilde{L}$ has at least 2^n states.

Section 5

Exercise 12. Compute the minimal ordered automaton and the syntactic ordered monoid of the languages of Exercise 8.

7 Notes

Only a few books cover the abstract definition of recognisable and rational subsets of a monoid: see Eilenberg [41], Berstel [13] or Sakarovitch [138]. The syntactic ordered monoid was first defined by Schützenberger [140].

Chapter V

Green's relations and local theory

Green's relations were introduced and studied by Green in 1951 [54]. These relations can be considered as a noncommutative generalisation to semigroups of the standard notion of being a multiple among integers or polynomials. They also have a natural interpretation in the Cayley graph of an A-generated monoid. They are essential for understanding the structure of semigroups.

1 Green's relations

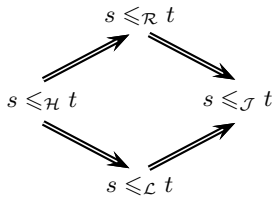
Let S be a semigroup. We define on S four preorder relations $\leqslant_{\mathcal{R}}, \leqslant_{\mathcal{L}}, \leqslant_{\mathcal{J}}$ and $\leqslant_{\mathcal{H}}$ as follows

$$
\begin{aligned}
s \leqslant_{\mathcal{R}} t &\text{ if and only if } s = tu \text{ for some } u \in S^1 \\
s \leqslant_{\mathcal{L}} t &\text{ if and only if } s = ut \text{ for some } u \in S^1 \\
s \leqslant_{\mathcal{J}} t &\text{ if and only if } s = utv \text{ for some } u, v \in S^1 \\
s \leqslant_{\mathcal{H}} t &\text{ if and only if } s \leqslant_{\mathcal{R}} t \text{ and } s \leqslant_{\mathcal{L}} t
\end{aligned}
$$

These relations can be considered as a noncommutative generalisation of the notion of multiple over the integers. For instance $s \leqslant_{\mathcal{R}} t$ if s is a *right multiple* of t, in the sense that one can pass from t to s by right multiplication by some element of S^1. These definitions can be reformulated in terms of ideals as follows

$$
\begin{aligned}
s \leqslant_{\mathcal{R}} t &\iff sS^1 \subseteq tS^1 \\
s \leqslant_{\mathcal{L}} t &\iff S^1 s \subseteq S^1 t \\
s \leqslant_{\mathcal{J}} t &\iff S^1 s S^1 \subseteq S^1 t S^1 \\
s \leqslant_{\mathcal{H}} t &\iff s \leqslant_{\mathcal{R}} t \text{ and } s \leqslant_{\mathcal{L}} t
\end{aligned}
$$

Thus $s \leqslant_{\mathcal{J}} t$ [$s \leqslant_{\mathcal{R}} t$, $s \leqslant_{\mathcal{L}} t$] if the ideal [right ideal, left ideal] generated by s is contained in the ideal [right ideal, left ideal] generated by t. The following diagram summarises the connections between these four preorders.



The equivalences associated with these four preorder relations are denoted by $\mathcal{R}$, $\mathcal{L}$, $\mathcal{J}$ and $\mathcal{H}$, respectively. Therefore

$$\begin{aligned} s \mathcal{R} t &\iff sS^1 = tS^1 \\ s \mathcal{L} t &\iff S^1 s = S^1 t \\ s \mathcal{J} t &\iff S^1 s S^1 = S^1 t S^1 \\ s \mathcal{H} t &\iff s \mathcal{R} t \text{ and } s \mathcal{L} t \end{aligned}$$

Thus two elements s and t are $\mathcal{R}$-equivalent if they generate the same right ideal, or, equivalently, if there exist $p, q \in S^1$ such that $s = tp$ and $t = sq$. The equivalence classes of the relation $\mathcal{R}$ are the *$\mathcal{R}$-classes* of S. The *$\mathcal{L}$-classes*, *$\mathcal{J}$-classes* and *$\mathcal{H}$-classes* are defined in a similar way. If s is an element of S, its $\mathcal{R}$-class [$\mathcal{L}$-class, $\mathcal{J}$-class, $\mathcal{H}$-class] is denoted by $R(s)$ [$L(s)$, $J(s)$, $H(s)$]. If $\mathcal{K}$ is one of the Green's relations, we shall use the notation $s <_{\mathcal{K}} t$ if $s \leqslant_{\mathcal{K}} t$ but $s \not\mathcal{K} t$.

If M is an A-generated monoid, the relations $\mathcal{R}$ and $\leqslant_{\mathcal{R}}$ [$\mathcal{L}$ and $\leqslant_{\mathcal{L}}$] have a natural interpretation on the right [left] Cayley graph of M.

Proposition 1.1. *Let M be an A-generated monoid and let $s, t \in M$. Then $s \leqslant_{\mathcal{R}} t$ [$s \leqslant_{\mathcal{L}} t$] if and only if there is a path from t to s in the right [left] Cayley graph of M. Furthermore $s \mathcal{R} t$ [$s \mathcal{L} t$] if and only if s and t are in the same strongly connected component of the right [left] Cayley graph of M.*

Proof. Suppose that $s \leqslant_{\mathcal{R}} t$. Then $tp = s$ for some $p \in M$. Since M is A-generated, p can be written as a product of elements of A, say $p = a_1 \cdots a_n$. Therefore there is a path

$$t \xrightarrow{a_1} ta_1 \xrightarrow{a_2} \cdots \quad t(a_1 \cdots a_{n-1}) \xrightarrow{a_n} t(a_1 \cdots a_n) = s$$

The other cases are similar. □

It follows in particular that the $\mathcal{R}$-classes [$\mathcal{L}$-classes] of M correspond to the strongly connected components of the right [left] Cayley graph.

Example 1.1. Let us consider the monoid M considered in Section IV.3.1.

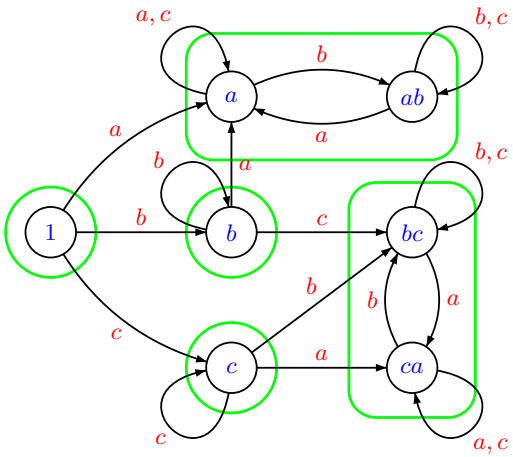


Figure 1.1. The right Cayley graph of M.

The right Cayley graph of M is represented in Figure 1.1. The strongly connected components of this graph are the $\mathcal{R}$-classes of M: $\{1\}$, $\{b\}$, $\{c\}$, $\{a, ab\}$ and $\{bc, ca\}$.

The next propositions summarise some useful properties of Green's relations.

Proposition 1.2. *In each semigroup S, the relations $\leqslant_{\mathcal{R}}$ and $\mathcal{R}$ are stable on the left and the relations $\leqslant_{\mathcal{L}}$ and $\mathcal{L}$ are stable on the right.*

Proof. Indeed, if $s \leqslant_{\mathcal{R}} t$, then $sS^1 \subseteq tS^1$ and thus $usS^1 \subseteq utS^1$. It follows that $us \leqslant_{\mathcal{R}} ut$. The other cases are analogous. □

Proposition 1.3. *Let S be a semigroup.*

(1) *Let e be an idempotent of S. Then $s \leqslant_{\mathcal{R}} e$ if and only if $es = s$ and $s \leqslant_{\mathcal{L}} e$ if and only if $se = s$.*

(2) *If $s \leqslant_{\mathcal{R}} sxy$, then $s \mathcal{R} sx \mathcal{R} sxy$. If $s \leqslant_{\mathcal{L}} yxs$, then $s \mathcal{L} xs \mathcal{L} yxs$.*

Proof. We shall prove only the first part of each statement, since the other part is dual.

(1) If $s \leqslant_{\mathcal{R}} e$, then $s = eu$ for some $u \in S^1$. It follows that $es = e(eu) = (ee)u = eu = s$. Conversely, if $es = s$, then $s \leqslant_{\mathcal{R}} e$ by definition.

(2) If $s \leqslant_{\mathcal{R}} sxy$, then $s \leqslant_{\mathcal{R}} sxy \leqslant_{\mathcal{R}} sx \leqslant_{\mathcal{R}} s$, whence $s \mathcal{R} sx \mathcal{R} sxy$. □

The first part of Proposition 1.3 can be extended to the preorder $\leqslant_{\mathcal{H}}$.

Proposition 1.4. *Let S be a semigroup. Let $s \in S$ and e be an idempotent of S. Then $s \leqslant_{\mathcal{H}} e$ if and only if $es = s = se$.*

Proof. It follows from Proposition 1.3. □

The restriction of the preorder $\leqslant_{\mathcal{H}}$ to $E(S)$ is actually an order, called the *natural order* on $E(S)$ and denoted by $\leqslant$.

Corollary 1.5. *Let S be a semigroup and let e and f be idempotents of S. The following conditions are equivalent:*

(1) $e \leqslant f$,

(2) $ef = e = fe$,
(3) $fef = e$.

Proof. The equivalence of (1) and (2) follows from Proposition 1.4 and that of (2) and (3) from the Simplification lemma. □

Despite its elementary nature, the next proposition is one of the cornerstones of semigroup theory.

Proposition 1.6. *In each semigroup S, the relations $\leqslant_{\mathcal{R}}$ and $\leqslant_{\mathcal{L}}$ [$\mathcal{R}$ and $\mathcal{L}$] commute.*

Proof. Suppose that $s \leqslant_{\mathcal{R}} r$ and $r \leqslant_{\mathcal{L}} t$. Then $s = rv$ and $r = ut$ for some $u, v \in S^1$. It follows that $s = utv \leqslant_{\mathcal{L}} tv \leqslant_{\mathcal{R}} t$. Thus $\leqslant_{\mathcal{L}} \circ \leqslant_{\mathcal{R}} \subseteq \leqslant_{\mathcal{R}} \circ \leqslant_{\mathcal{L}}$. The opposite inclusion holds by duality and hence $\leqslant_{\mathcal{R}}$ and $\leqslant_{\mathcal{L}}$ commute. The proof for $\mathcal{R}$ and $\mathcal{L}$ is similar. □

Here is a first consequence of Proposition 1.6.

Proposition 1.7. *The relation $\leqslant_{\mathcal{J}}$ is equal to $\leqslant_{\mathcal{L}} \circ \leqslant_{\mathcal{R}}$ and to $\leqslant_{\mathcal{R}} \circ \leqslant_{\mathcal{L}}$. It is also the preorder generated by $\leqslant_{\mathcal{R}}$ and $\leqslant_{\mathcal{L}}$.*

Proof. It follows directly from Proposition I.2.21. □

We now introduce the fifth Green's relation $\mathcal{D}$, which is the preorder generated by $\mathcal{R}$ and $\mathcal{L}$. Propositions 1.6 and I.2.21 immediately leads to an easier definition.

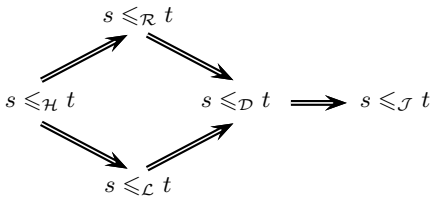
Proposition 1.8. *The relation $\mathcal{D}$ is equal to $\mathcal{L} \circ \mathcal{R}$ and to $\mathcal{R} \circ \mathcal{L}$.*

One can therefore give the following definition of $\mathcal{D}$:

$$\begin{aligned} s \mathcal{D} t &\iff \text{there exists } u \in S \text{ such that } s \mathcal{R} u \text{ and } u \mathcal{L} t \\ &\iff \text{there exists } v \in S \text{ such that } s \mathcal{L} v \text{ and } v \mathcal{R} t. \end{aligned}$$

The equivalence classes of $\mathcal{D}$ are called the $\mathcal{D}$-*classes* of S, and the $\mathcal{D}$-class of an element s is denoted by $D(s)$.

It is easy to see that $s \mathcal{D} t$ implies $s \mathcal{J} t$. Indeed, if $s \mathcal{D} t$, there exists $u \in S$ such that $s \mathcal{R} u$ and $u \mathcal{L} t$. It follows that $sS^1 = uS^1$ and $S^1u = S^1t$, whence $S^1sS^1 = S^1tS^1$. The following diagram summarises the connections between the five Green's relations.



The equality $\mathcal{D} = \mathcal{J}$ does not hold in arbitrary semigroups (see Example 1.2 below) but it holds for finite semigroups.

Theorem 1.9. *In a finite semigroup, the Green's relations $\mathcal{J}$ and $\mathcal{D}$ are equal. Furthermore, the following properties hold:*

(1) *If* $s \leqslant_{\mathcal{J}} sx$ *(in particular if* $s \mathcal{J} sx$*), then* $s \mathcal{R} sx$*;*
(2) *If* $s \leqslant_{\mathcal{J}} xs$ *(in particular if* $s \mathcal{J} xs$*), then* $s \mathcal{L} xs$*.*
(3) *If* $s \mathcal{J} t$ *and* $s \leqslant_{\mathcal{R}} t$*, then* $s \mathcal{R} t$*;*
(4) *If* $s \mathcal{J} t$ *and* $s \leqslant_{\mathcal{L}} t$*, then* $s \mathcal{L} t$*;*
(5) *if* $s = usv$ *for some* $u, v \in S^1$*, then* $us \mathcal{H} s \mathcal{H} sv$*.*

Proof. If $x \mathcal{D} y$, there exist $z \in S$ such that $x \mathcal{R} z$ and $z \mathcal{L} y$. It follows that $x \mathcal{J} z$ and $z \mathcal{J} y$, whence $x \mathcal{J} y$.

Conversely, suppose that $x \mathcal{J} y$. Then there exist $s, t, u, v \in S^1$ such that $y = txu$ and $x = syv$, whence $x = stxuv$. By multiplying on the left by st and on the right by uv, we obtain by induction the relation $(st)^n x(uv)^n = x$ for all $n > 0$. By Proposition II.6.33, one can choose n such that both $(st)^n$ and $(uv)^n$ are idempotent. It follows that $(st)^n x = (st)^n (st)^n x(uv)^n = (st)^n x(uv)^n = x$ and similarly $x = x(uv)^n$. Therefore $tx \mathcal{L} x$ and $xu \mathcal{R} x$. The first relation implies $y = txu \mathcal{L} xu$ and finally $y \mathcal{D} x$.

(1) If $s \leqslant_{\mathcal{J}} sx$, there exist $u, v \in S^1$ such that $usxv = s$. By multiplying on the left by u and on the right by xv, we obtain by induction the relation $u^n s(xv)^n = s$ for all $n > 0$. By Proposition II.6.33, one can choose n such that u^n is idempotent. It follows that $s = u^n s(xv)^n = u^n u^n s(xv)^n = u^n s$, whence $s(xv)^n = s$. It follows that $s \mathcal{R} sx$, since $(sx)(v(xv)^{n-1}) = s$.

(2) is dual from (1).

(3) If $s \leqslant_{\mathcal{R}} t$, there exist $u \in S^1$ such that $s = tu$. If further $s \mathcal{J} t$, then $t \mathcal{J} tu$ and $t \mathcal{R} tu$ by (1). Thus $s \mathcal{R} t$.

(4) is dual from (3).

(5) If $s = usv$ then $s \leqslant_{\mathcal{J}} us$. It follows by (1) that $s \mathcal{R} sv$ and a dual argument shows that $s \mathcal{L} us$. Since the relation $\mathcal{R}$ is stable on the left, one has $us \mathcal{R} usv = s$ and dually, $sv \mathcal{L} s$. Thus $us \mathcal{H} s \mathcal{H} sv$. □

Example 1.2. Let S be the infinite semigroup of matrices of the form

$$\begin{pmatrix} a & 0 \\ b & 1 \end{pmatrix}$$

where a and b are strictly positive rational numbers, equipped with the usual multiplication of matrices. Then the four relations $\mathcal{R}$, $\mathcal{L}$, $\mathcal{H}$ and $\mathcal{D}$ coincide with the equality, but S has a single $\mathcal{J}$-class.

Proposition 1.8 shows that for two elements s and t, the three conditions $s \mathcal{D} t$, $R(s) \cap L(t) \neq \emptyset$ and $L(s) \cap R(t) \neq \emptyset$ are equivalent. It is therefore possible to represent $\mathcal{D}$-classes by an "egg-box picture", as in Figure 1.2.

*					*
		*			
				*	
	*		*		

Figure 1.2. A $\mathcal{D}$-class containing some idempotents.

Each row represents an $\mathcal{R}$-class, each column an $\mathcal{L}$-class and each cell an $\mathcal{H}$-class. The possible presence of an idempotent within an $\mathcal{H}$-class is indicated by a star. We shall see later (Proposition 1.13) that these $\mathcal{H}$-classes containing an idempotent are groups, and that all such groups occurring within a given $\mathcal{D}$-class are isomorphic. The next proposition describes the structure of a $\mathcal{D}$-class.

Proposition 1.10 (Green's lemma)**.** *Let s and t be two $\mathcal{R}$-equivalent elements of S. If $s = tp$ and $t = sq$ for some $p, q \in S^1$, then the map $x \to xp$ is a bijection from $L(t)$ onto $L(s)$, the map $x \to xq$ is a bijection from $L(s)$ and $L(t)$. Moreover, these bijections preserve the $\mathcal{H}$-classes and are inverse one from the other.*

Proof. Let $n \in L(s)$ (see Figure 1.3). Since $\mathcal{L}$ is stable on the right, $nq \in L(sq)$. Furthermore, there exist $u \in S^1$ such that $n = us$, whence $nqp = usqp = utp = us = n$. Similarly, if $m \in L(t)$, then $mpq = m$ and thus the maps $x \to xp$ and $x \to xq$ define inverse bijections between $L(s)$ and $L(t)$. Moreover, Proposition 1.2 shows that the maps $x \to xp$ and $x \to xq$ preserve the $\mathcal{H}$-classes. □

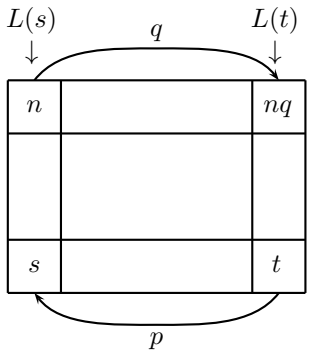


Figure 1.3. An illustration of Green's lemma.

There is of course a dual version of Green's lemma for $\mathcal{L}$-equivalent elements. Green's lemma has several important consequences. First, the "Location Theorem" of Clifford and Miller:

Theorem 1.11 (Location Theorem)**.** *Let D be a $\mathcal{D}$-class of a semigroup S, and let s and t be elements of D. The following conditions are equivalent:*

(1) *$st \in R(s) \cap L(t)$,*

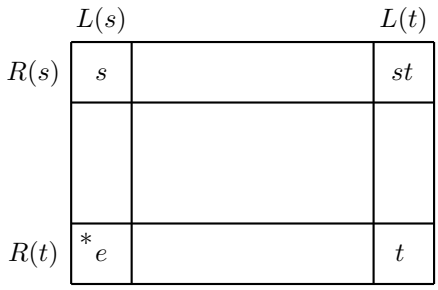
(2) *$R(t) \cap L(s)$ contains an idempotent,*

(3) *$\bar{s}st = t$ and $st\bar{t} = s$ for some inverse $\bar{s}$ of s and some inverse $\bar{t}$ of t.*

If S is finite, then these conditions are equivalent to

(4) *$st \in D$.*

Proof. (1) implies (2). If $st \in R(s) \cap L(t)$, the multiplication on the right by t is, by Green's lemma, a bijection from $L(s)$ onto $L(t)$ preserving the $\mathcal{H}$-classes. In particular, there exists an element $e \in R(t) \cap L(s)$ such that $et = t$. Since $e \mathcal{R} t$, one has $e = tv$ for some $v \in S^1$ and $ee = etv = tv = e$. Thus e is idempotent.



(2) implies (3). Let e be an idempotent of $R(t) \cap L(s)$. First, one gets $et = t$ and $se = s$ by Proposition 1.3. Since $t \mathcal{R} e$, $e = tt'$ for some $t' \in S^1$. Setting $\bar{t} = t'e$, we get $t\bar{t}t = tt'et = eet = t$ and $\bar{t}t\bar{t} = t'ett'e = t'e = \bar{t}$. Thus $\bar{t}$ is an inverse of t. Furthermore, $st\bar{t} = stt'e = se = s$. The proof of the existence of $\bar{s}$ is dual.
(3) implies (1) is clear.
Finally, if S is finite, Conditions (1) and (4) are equivalent By Theorem 1.9. □

Proposition 1.12. *Let D be a $\mathcal{D}$-class of a semigroup S. If D contains an idempotent, it contains at least one idempotent in each $\mathcal{R}$-class and in each $\mathcal{L}$-class.*

Proof. Suppose that D contains an idempotent e and let $s \in D$. Then $e \mathcal{R} r$ and $r \mathcal{L} s$ for some $r \in D$. Thus $er = r$ by Proposition 1.3 and $ru = e$ for some $u \in S^1$. It follows that ur is idempotent, since $urur = u(ru)r = uer = ur$. Furthermore $r(ur) = er = r$. Consequently $r \mathcal{L} ur$, $L(s) = L(r) = L(ur)$ and thus the $\mathcal{L}$-class of s contains an idempotent. □

Here is a useful consequence of the Location Theorem.

Proposition 1.13. *Let H be an $\mathcal{H}$-class of a semigroup S. The following conditions are equivalent:*

(1) *H contains an idempotent,*
(2) *there exist $s, t \in H$ such that $st \in H$,*
(3) *H is a group.*

Moreover, every group of S is contained in an $\mathcal{H}$-class.

Proof. The equivalence of (1) and (2) follows from Theorem 1.11. Furthermore, it is clear that (3) implies (1). Let us show that (1) implies (3).

Let H be a $\mathcal{H}$-class containing an idempotent e. Then H is a semigroup: indeed, if $s, t \in H$, we have $st \in R(s) \cap L(t) = H$. Moreover, if $s \in H$, we have $s \mathcal{R} e$ and hence $es = s$ by Proposition 1.3. Moreover, by Green's lemma, the map $x \to xs$ is a permutation on H. In particular, there exists $t \in H$ such that $ts = e$, and thus H is a group with identity e.

Finally, if H is a group with identity e, then H is contained in the $\mathcal{H}$-class of e. Indeed, if t is the inverse of an element s of H, then $st = ts = e$ and $se = es = s$, which proves that $s \mathcal{H} e$. □

The following is another remarkable consequence of Green's lemma.

Proposition 1.14. *Two maximal subgroups of a $\mathcal{D}$-class are isomorphic.*

Proof. From Proposition 1.13, the two groups are of the form $H(e)$ and $H(f)$ for some idempotent e, f of the same $\mathcal{D}$-class D. Since $e \mathcal{D} f$, there exists $s \in R(e) \cap L(f)$. Thus $es = s$, $sf = s$ and $ts = f$ for some $t \in S^1$. By Green's lemma, the function φ defined by $\varphi(x) = txs$ is a bijection from $H(e)$ onto $H(f)$, which maps e to f since $tes = ts = f$.

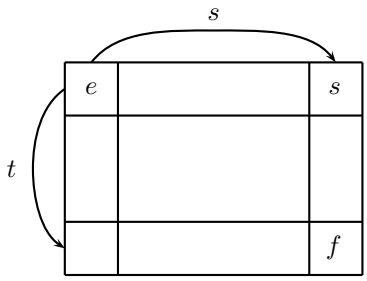


Figure 1.4. The $\mathcal{D}$-class D.

We claim that φ is a morphism. First observe that st is idempotent, since $ststst = sftst = stst$. Furthermore, $st \mathcal{R} s$ since $sts = sf = s$. If $y \in H(e)$, then $y \mathcal{R} e \mathcal{R} s \mathcal{R} st$ and by Proposition 1.3, $(st)y = y$. It follows that for all $x, y \in H(e)$,

$$\varphi(xy) = txys = tx(sty)s = (txs)(tys) = \varphi(x)\varphi(y)$$

proving the claim. Thus $H(e)$ and $H(f)$ are isomorphic. □

For finite semigroups, some further properties hold.

Corollary 1.15. *Let S be a finite semigroup and let $s, t \in S$ be two $\mathcal{J}$-related elements of S. If $st \mathcal{L} s$ or $ts \mathcal{R} s$, then $H(t)$ is a group. If $st = s$ or $ts = s$, then t is idempotent.*

Proof. Suppose that $st \mathcal{L} s$ (the other case is dual) and let J be the common $\mathcal{J}$-class of s, t and st. Since $st \leqslant_{\mathcal{L}} t$, Theorem 1.9 shows that $t \mathcal{L} st$, whence $s \mathcal{L} t$ since $st \mathcal{L} s$. Thus $L(s) = L(t)$ and $R(t) \cap L(s) = H(t)$. Since $st \in J$, Theorem 1.11 shows that $H(t)$ contains an idempotent. Thus by Proposition 1.13, $H(t)$ is a group.

Suppose that $st = s$ and let e be the idempotent of $H(t)$. By Green's lemma, the left multiplication by s induces a bijection from $H(t)$ onto $H(st)$. But since $e \mathcal{L} s$, $se = s$ by Proposition 1.3. Thus $se = s = st$, whence $e = t$. □

The case $t = s$ is worth a separate statement, that should be compared with Proposition 1.13.

Corollary 1.16. *Let S be a finite semigroup and let $s \in S$. If $s \mathcal{J} s^2$, then $H(s)$ is a group.*

We conclude this section by two results on the maximal $\mathcal{J}$-classes of a finite semigroup.

Proposition 1.17. *In a finite monoid, the $\mathcal{J}$-class of the identity is a group.*

Proof. Let J be the $\mathcal{J}$-class of 1. If e is an idempotent of J, then $e \leqslant 1$ by Corollary 1.5 whence $e = 1$ by Theorem 1.9. It follows that J contains a unique idempotent and by Proposition 1.12, a unique $\mathcal{R}$-class and a unique $\mathcal{L}$-class. Thus J is an $\mathcal{H}$-class and thus a group. □

The proof of the second result requires a forward reference to Corollary 2.25, but it fits better in this section than in the next one.

Proposition 1.18. *Let J be a $\leqslant_{\mathcal{J}}$-maximal $\mathcal{J}$-class of a finite semigroup S. Then J is either regular or reduced to a single null element.*

Proof. Let J be a maximal $\mathcal{J}$-class. Suppose that J contains two distinct elements s and t. Then $s = utv$ and $t = xsy$ for some $x, y, u, v \in S^1$. Thus $s = uxsyv$ and since $s \neq t$, we may assume that $(u, v) \neq (1, 1)$ whence $u \in S$ or $v \in S$. Suppose that $u \in S$, the other case being dual. Then $ux \in S$ and since $s \leqslant_{\mathcal{J}} ux$ and J is maximal, it follows that $s \mathcal{J} ux$. Similarly, $s \leqslant_{\mathcal{J}} svy$, whence $svy \in J$. Thus J contains ux, svy and their product. Therefore it is regular by Corollary 2.25. □

2 Inverses, weak inverses and regular elements

In this section, we study in more detail the notion of a semigroup inverse introduced in Chapter II.

2.1 Inverses and weak inverses

An element $\bar{s}$ of a semigroup S is a *weak inverse* of an element s if $\bar{s}s\bar{s} = \bar{s}$. It is an *inverse* (or a *semigroup inverse*) of s if, furthermore, $s\bar{s}s = s$. Note that any idempotent is its own inverse.

We let $W(s)$ $[V(s)]$ denote the set of all weak inverses [inverses] of the element s.

Proposition 2.19. *If $\bar{s}$ is a weak inverse of s, then $\bar{s}s$ and $s\bar{s}$ are idempotents and $s\bar{s}s$ is an inverse of $\bar{s}$. Furthermore, the relations $s\bar{s}$ $\mathcal{L}$ $\bar{s}$ $\mathcal{R}$ $\bar{s}s$ and $\bar{s}s$ $\mathcal{L}$ $s\bar{s}s$ $\mathcal{R}$ $s\bar{s}$ hold.*

Proof. If $\bar{s}$ is a weak inverse of s, then $\bar{s}s\bar{s} = \bar{s}$. Thus $s\bar{s}s\bar{s} = s\bar{s}$ and $\bar{s}s\bar{s}s = \bar{s}s$. Moreover, since $\bar{s}(s\bar{s}s)\bar{s} = \bar{s}s\bar{s}$ and $(s\bar{s}s)\bar{s}(s\bar{s}s) = s\bar{s}s$, $s\bar{s}s$ is an inverse of $\bar{s}$. The relations of the statement follow immediately. □

Corollary 2.20. *If $\bar{s}$ is an inverse of s, then $\bar{s}s$ and $s\bar{s}$ are idempotents. Furthermore, s and $\bar{s}$ are in the same $\mathcal{D}$-class and the relations $s\bar{s}$ $\mathcal{L}$ $\bar{s}$ $\mathcal{R}$ $\bar{s}s$ $\mathcal{L}$ s $\mathcal{R}$ $s\bar{s}$ hold.*

Proof. If $\bar{s}$ is an inverse of s, then $s\bar{s}s = s$ and the result follows from Proposition 2.19. □

The case where $\bar{s}$ is a weak inverse [an inverse] of s is depicted in Figure 2.1. However, it may happen that some of the elements represented in different $\mathcal{H}$-classes are actually in the same $\mathcal{H}$-class. In particular, if $s = \bar{s}$, the $\mathcal{H}$-class of s is a group H whose identity is the idempotent $e = s\bar{s} = \bar{s}s$. Furthermore s, $\bar{s}$ and e are all in H and $\bar{s}$ is the group inverse of s in H.

Figure 2.1. Two egg-box pictures. On the left, $\bar{s}$ is a weak inverse of s. On the right, $\bar{s}$ is an inverse of s.

In general, an element may have several inverses. However, it has at most one inverse in a given $\mathcal{H}$-class.

Proposition 2.21. *An $\mathcal{H}$-class H contains an inverse $\bar{s}$ of an element s if and only if $R(H) \cap L(s)$ and $R(s) \cap L(H)$ contain an idempotent. In this case, H contains a unique inverse of s.*

Proof. Suppose that H contains an inverse $\bar{s}$ of s. Then by Corollary 2.20, the idempotent $\bar{s}s$ belongs to $R(\bar{s}) \cap L(s)$ and the idempotent $s\bar{s}$ to $R(s) \cap L(\bar{s})$. Conversely, suppose that $R(s) \cap L(H)$ contains an idempotent e and $R(H) \cap L(s)$ an idempotent f. Then $e \mathcal{R} s$ and thus $es = s$ by Proposition 1.3. Now, by Green's lemma, there exists a unique element $\bar{s} \in H$ such that $\bar{s}s = f$. Since $s \mathcal{L} f$, $sf = s$ and hence $s\bar{s}s = s$. Similarly, $f \mathcal{R} \bar{s}$, whence $f\bar{s} = \bar{s}$ and $\bar{s}s\bar{s} = \bar{s}$. Thus $\bar{s}$ is an inverse of s.

Finally, suppose that H contains two inverses $\bar{s}_1$ and $\bar{s}_2$ of s. Then Corollary 2.20 shows that $s\bar{s}_1$ and $s\bar{s}_2$ are idempotents of the same $\mathcal{H}$-class and hence are equal. Similarly, $\bar{s}_1 s$ and $\bar{s}_2 s$ are equal. It follows that $\bar{s}_1 s \bar{s}_1 = \bar{s}_2 s \bar{s}_1 = \bar{s}_2 s \bar{s}_2$, that is, $\bar{s}_1 = \bar{s}_2$. □

Two elements s and t of a semigroup S are said to be *conjugate* if there exist $u, v \in S^1$ such that $s = uv$ and $t = vu$. Conjugate idempotents can be characterised as follows:

Proposition 2.22. *Let e and f be two idempotents of a semigroup S. The following conditions are equivalent:*

(1) *e and f are conjugate,*

(2) *there exist two elements $u, v \in S$ such that $u \mathcal{D} v \mathcal{D} e \mathcal{D} f$, $uv = e$ and $vu = f$,*

(3) *e and f are $\mathcal{D}$-equivalent.*

Proof. It is clear that (2) implies (1) and (3).

(1) implies (3). Suppose first that $e = uv$ and $f = vu$ for some $u, v \in S^1$. Then $uvuv = uv$ and $vuvu = vu$, whence $uv \mathcal{R} uvu$ and $uvu \mathcal{L} vu$. Thus $e = uv \mathcal{D} vu = f$.

(3) implies (2). Suppose that $e \mathcal{D} f$. Then there exists $s \in S$ such that $e \mathcal{R} s$ and $s \mathcal{L} f$. By Green's lemma, there exists an element $\bar{s} \in L(e) \cap R(f)$ such that $\bar{s}s = f$. Thus $s\bar{s}s = sf = s$ and $\bar{s}s\bar{s} = f\bar{s} = \bar{s}$. It follows that $\bar{s}$ is an inverse of s. By Corollary 2.20, $s\bar{s}$ is an idempotent of the same $\mathcal{H}$-class as e and thus is equal to e. □

We conclude this section by an elementary result on idempotents.

Proposition 2.23. *Let e be an idempotent of a semigroup S. If $e = xy$ for some $x, y \in S$, then ex and ye are mutually inverse elements.*

Proof. Indeed, $(ex)(ye)(ex) = exyex = ex$ and $(ye)(ex)(ye) = yexye = ye$. □

2.2 Regular elements

An element is *regular* if it has at least one inverse. A semigroup is called *regular* if all its elements are regular. Similarly, a $\mathcal{D}$-class [$\mathcal{L}$-class, $\mathcal{R}$-class, $\mathcal{J}$-class] is called *regular* if all its elements are regular. A nonregular $\mathcal{D}$-class is also called a *null* $\mathcal{D}$-class.

The set of regular elements of a semigroup S is denoted by $\mathrm{Reg}(S)$. Since an idempotent is its own inverse, $E(S)$ is a subset of $\mathrm{Reg}(S)$.

The next proposition gives various characterisations of regular elements.

Proposition 2.24. *Let s be an element of a semigroup S. The following conditions are equivalent:*

(1) *s is regular,*
(2) *$s\bar{s}s = s$ for some $\bar{s} \in S$,*
(3) *$D(s)$ contains an idempotent,*
(4) *$R(s)$ contains an idempotent,*
(5) *$L(s)$ contains an idempotent.*

Proof. (1) implies (2) by definition. Condition (2) states that s is a weak inverse of $\bar{s}$. Thus Proposition 2.19 shows that (2) implies (1), (3), (4) and (5). The equivalence of (3), (4) and (5) follows from Proposition 1.12.

(4) implies (1). Let e be an idempotent such that $s \mathcal{R} e$. Then $es = s$ and $st = e$ for some $t \in S^1$. We claim that $\bar{s} = tst$ is an inverse of s. Indeed $s\bar{s}s = ststs = ees = es = s$ and $\bar{s}s\bar{s} = tststst = \bar{s}$. Thus s is regular. □

It is useful to restate Proposition 2.24 in terms of $\mathcal{D}$-classes.

Corollary 2.25. *Let $\mathcal{D}$ be a $\mathcal{D}$-class of a finite semigroup. The following conditions are equivalent:*

(1) *D is regular,*
(2) *D contains a regular element,*
(3) *D contains an idempotent,*
(4) *each $\mathcal{R}$-class of D contains an idempotent,*
(5) *each $\mathcal{L}$-class of D contains an idempotent,*
(6) *there exist two elements of D whose product belongs to D.*

Corollary 2.25 shows that a regular $\mathcal{D}$-class contains at least one idempotent in each $\mathcal{R}$-class and in each $\mathcal{L}$-class. It follows that all the $\mathcal{R}$-classes and $\mathcal{L}$-classes contained in a regular $\mathcal{D}$-class are regular.

Let $\mathcal{K}$ be one of the Green's relations $\mathcal{R}$, $\mathcal{L}$, $\mathcal{J}$ or $\mathcal{H}$. A semigroup is *$\mathcal{K}$-trivial* if and only if $a \mathcal{K} b$ implies $a = b$.

Proposition 2.26. *Let S be a finite semigroup and let $\mathcal{K}$ be one of the Green's relations $\mathcal{R}$, $\mathcal{L}$, $\mathcal{H}$ or $\mathcal{J}$. Then S is $\mathcal{K}$-trivial if and only if its regular $\mathcal{K}$-classes are trivial.*

Proof. Let n be the exponent of S.

(a) $\mathcal{K} = \mathcal{R}$. Suppose $x \mathcal{R} y$. Then there exist $c, d \in S^1$ such that $xc = y$, $yd = x$, whence $xcd = x$. One has $(cd)^n \mathcal{R} (cd)^n c$ and since the restriction of $\mathcal{R}$ to the regular $\mathcal{R}$-class is trivial, one gets $(cd)^n = (cd)^n c$. It follows that $x = x(cd)^n = x(cd)^n c = xc = y$ and therefore S is $\mathcal{R}$-trivial.

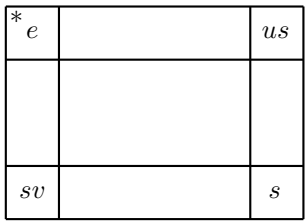
(b) $\mathcal{K} = \mathcal{L}$. The proof is dual.

(c) $\mathcal{K} = \mathcal{J}$. By (a) and (b), S is $\mathcal{R}$-trivial and $\mathcal{L}$-trivial. Since $\mathcal{J} = \mathcal{D} = \mathcal{R} \circ \mathcal{L}$, S is $\mathcal{J}$-trivial.

(d) $\mathcal{K} = \mathcal{H}$. Suppose that $x \mathcal{H} y$. Then there exist $a, b, c, d \in S^1$ such that $ax = y$, $by = x$, $xc = y$ and $yd = x$. It follows that $x = axd$ and therefore $a^n x d^n = x$. Since a^n is idempotent and since $a^{n+1} \mathcal{H} a^n$, one gets $a^{n+1} = a^n$ since the $\mathcal{H}$-class of a^n is regular. It follows that $x = a^n x d^n = a^{n+1} x d^n = a(a^n x d^n) = ax = y$. Therefore S is $\mathcal{H}$-trivial. □

A finite $\mathcal{H}$-trivial semigroup is also called *aperiodic*. See Proposition VI.2.1 for more details. We conclude this section by another property of finite semigroups.

Proposition 2.27. *Let S be a finite semigroup and let T be a subsemigroup of S. Let $s \in T$ and let e be an idempotent of S. Suppose that, for some $u, v \in T^1$, $e \mathcal{R}_S us$, $us \mathcal{L}_S s$, $s \mathcal{R}_S sv$ and $sv \mathcal{L}_S e$. Then $e \in T$ and $e \mathcal{R}_T us \mathcal{L}_T s \mathcal{R}_T sv \mathcal{L}_T e$.*



Proof. Since $R_S(us) \cap L_S(sv)$ contains an idempotent, Green's lemma shows that $svus$ belongs to $R_S(sv) \cap L_S(us)$, which is equal to $H_S(s)$. It follows that the right translation ρ_{vus} is a permutation on $H_S(s)$. Since T is finite, some power of vus, say $(vus)^n$ is an idempotent f of T. Since ρ_{vus} is a permutation on $H_S(s)$, ρ_f is also a permutation on $H_S(s)$ and since $f^2 = f$, this permutation is the identity. In particular, $sf = s$, that is, $s(vus)^n = s$. It follows that $s \mathcal{R}_T sv \mathcal{R}_T svus$ and a dual argument shows that $s \mathcal{L}_T us \mathcal{L}_T svus$. Thus $svus \in R_T(sv) \cap L_T(us)$ and by Green's lemma again, $R_T(us) \cap L_T(sv)$ contains an idempotent. This idempotent belongs to the $\mathcal{H}$-class $R_S(us) \cap L_S(sv)$, thereby it is equal to e. Thus $e \in T$. □

3 Rees matrix semigroups

The Location Theorem indicates that the product of two elements s and t of the same $\mathcal{D}$-class D either falls out of D or belongs to the intersection of the $\mathcal{R}$-class of s and of the $\mathcal{L}$-class of t. In the latter case, the location of st in the egg-box picture is precisely known and the intersection of the $\mathcal{R}$-class of t and of the $\mathcal{L}$-class of s is a group. This suggests that the structure of a regular $\mathcal{D}$-class depends primarily on the coordinates of its elements in the egg-box picture and on its maximal groups. This motivates the following definition.

Let I and J be two nonempty sets, G be a group and $P = (p_{j,i})_{j \in J, i \in I}$ be a $J \times I$-matrix with entries in G. The *Rees matrix semigroup* with G as *structure group*, P as *sandwich matrix* and I and J as indexing sets, is the semigroup

$M(G, I, J, P)$ defined on the set $I \times G \times J$ by the operation

$$(i, g, j)(i', g', j') = (i, gp_{j,i'}g', j') \tag{3.1}$$

More generally, if $P = (p_{j,i})_{j\in J, i\in I}$ is a $J \times I$-matrix with entries in G^0, we let $M^0(G, I, J, P)$ denote the semigroup, called a *Rees matrix semigroup with zero*, defined on the set $(I \times G \times J) \cup 0$ by the operation

$$(i, g, j)(i', g', j') = \begin{cases} (i, gp_{j,i'}g', j') & \text{if } p_{j,i'} \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

Example 3.1. A *Brandt semigroup* is a Rees matrix semigroup in which $I = J$ and P is the identity matrix. Therefore, the product is defined by

$$(i, g, j)(i', g', j') = \begin{cases} (i, gg', j') & \text{if } j = i', \\ 0 & \text{otherwise.} \end{cases}$$

A *Brandt aperiodic semigroup* is a Brandt semigroup whose structure group is trivial. If $I = \{1, \ldots, n\}$, this semigroup is denoted by B_n. For instance, B_2 is the semigroup of 2×2 Boolean matrices

$$B_2 = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \right\}$$

under multiplication. This semigroup is studied in more detail in Section II.2.1.

The regularity of a Rees matrix semigroups depends only on its sandwich matrix.

Proposition 3.28. *A Rees matrix semigroup with zero is regular if and only if every row and every column of its sandwich matrix has a nonzero entry.*

Proof. Let $S = M^0(G, I, J, P)$ be a regular Rees matrix semigroup with zero. Let $i \in I$, $j \in J$ and let $s = (i, g, j)$ be a nonzero element of S. Since s is regular, there exists a nonzero element $\bar{s} = (i', g', j')$ such that $s\bar{s}s = s$. It follows that $p_{j,i'} \neq 0$ and $p_{j',i} \neq 0$. Since this holds for every $i \in I$ and $j \in J$, each row and each column of P has a nonzero entry.

Conversely, assume that in P, every row j contains a nonzero entry p_{j,i_j} and every column i contains a nonzero entry $p_{j_i,i}$. Then each nonzero element $s = (i, g, j)$ admits as an inverse the element $\bar{s} = (i_j, p_{j,i_j}^{-1}g^{-1}p_{j_i,i}^{-1}, j_i)$ since

$$\begin{aligned} s\bar{s}s &= (i, gp_{j,i_j}p_{j,i_j}^{-1}g^{-1}p_{j_i,i}^{-1}p_{j_i,i}g, j) = s \text{ and} \\ \bar{s}s\bar{s} &= (i_j, p_{j,i_j}^{-1}g^{-1}p_{j_i,i}^{-1}p_{j_i,i}gp_{j,i_j}p_{j,i_j}^{-1}g^{-1}p_{j_i,i}^{-1}, j_i) = \bar{s} \end{aligned}$$

Thus S is regular. □

Green's relations in a regular Rees matrix semigroup with zero are easy to describe. We remind the reader that a semigroup S is simple if its only ideals are $\emptyset$ and S. It is 0-simple if it has a zero, denoted by 0, if $S^2 \neq 0$ and if $\emptyset$, 0 and S are its only ideals. In a semigroup with zero, an idempotent e is said to be 0-*minimal* if $e \neq 0$ and if an idempotent f satysfying $0 \leqslant f \leqslant e$ is either equal to 0 or to e. Here $\leqslant$ refers to the natural order on idempotent defined on page 101.

Proposition 3.29. *Let $S = M^0(G, I, J, P)$ be a regular Rees matrix semigroup with zero. Then S is 0-simple. In particular, $\mathcal{D} = \mathcal{J}$ in S and all the elements of $S - 0$ are in the same $\mathcal{D}$-class. Furthermore, if $s = (i, g, j)$ and $s' = (i', g', j')$ are two elements of $S - 0$, then*

$$s \leqslant_{\mathcal{R}} s' \iff s \mathcal{R} s' \iff i = i', \tag{3.2}$$

$$s \leqslant_{\mathcal{L}} s' \iff s \mathcal{L} s' \iff j = j', \tag{3.3}$$

$$s \leqslant_{\mathcal{H}} s' \iff s \mathcal{H} s' \iff i = i' \text{ and } j = j'. \tag{3.4}$$

Proof. Proposition 3.28 implies that in P, every row j contains a nonzero entry p_{j,i_j} and every column i contains a nonzero entry $p_{j_i,i}$.

Formula 3.1 shows that if $s \leqslant_{\mathcal{R}} s'$, then $i = i'$. The converse is true, since

$$(i, g, j)(i_j, p_{j,i_j}^{-1} g^{-1} g', j') = (i, g p_{j,i_j} p_{j,i_j}^{-1} g^{-1} g', j') = (i, g', j')$$

This proves (3.2). Property (3.3) is dual and (3.4) is the conjunction of (3.2) and (3.3).

Setting $t = (i, 1, j')$, it follows in particular that $s \mathcal{R} t \mathcal{L} s'$, whence $s \mathcal{D} s'$. Thus the relations $\mathcal{D}$ and hence $\mathcal{J}$ are universal on $S - 0$. Finally, if e and f are nonzero idempotents such that $e \leqslant f$ (that is, $e \leqslant_{\mathcal{H}} f$), then $e \mathcal{H} f$ by (3.4), and hence $e = f$ by Proposition 1.13. Thus every nonzero idempotent of S is 0-minimal and S is 0-simple. □

	j		j'
i	(i, g, j)		$(i, g p_{j,i'} g', j')$
i'	$*$		(i', g', j')

Figure 3.1. The product of two elements when $p_{j,i'} \neq 0$.

An analogous result holds for Rees matrix semigroups.

Proposition 3.30. *Let $S = M(G, I, J, P)$ be a Rees matrix semigroup. Then S is simple. In particular, $\mathcal{D}$ and $\mathcal{J}$ are both the universal relation and every $\mathcal{H}$-class is a group. Furthermore, if $s = (i, g, j)$ and $s' = (i', g', j')$ are two elements of S, then*

$$s \leqslant_{\mathcal{R}} s' \iff s \mathcal{R} s' \iff i = i', \tag{3.5}$$

$$s \leqslant_{\mathcal{L}} s' \iff s \mathcal{L} s' \iff j = j', \tag{3.6}$$

$$s \leqslant_{\mathcal{H}} s' \iff s \mathcal{H} s' \iff i = i' \text{ and } j = j'. \tag{3.7}$$

Proof. The proposition mostly follows from Proposition 3.29 by considering S^0. A complementary property is that $s \mathcal{R} ss' \mathcal{L} s'$, which shows that the relations $\mathcal{D}$ and $\mathcal{J}$ are universal on S. It follows that S is simple. Taking $s = s'$, we get $s \mathcal{H} s^2$ and thus by Proposition 1.13, $H(s)$ is a group. Consequently, each $\mathcal{H}$-class is a group. □

By Proposition II.3.19, a simple semigroup has a single $\mathcal{J}$-class, and a 0-simple semigroup has a single nonzero $\mathcal{J}$-class.

Proposition 3.31. *A finite 0-simple semigroup contains a single nonzero $\mathcal{D}$-class and this $\mathcal{D}$-class is regular.*

Proof. Let S^0 be a simple semigroup. By Proposition II.3.19, S is a $\mathcal{D}$-class of S^0. Furthermore, if $s \in S$, then $s^2 \in S$ and Corollary 1.16 shows that $s \mathcal{H} s^2$. It follows by Proposition 1.13 that $H(s)$ is a group and thus S is a regular $\mathcal{D}$-class of S^0. □

A similar result holds for finite simple semigroups.

Proposition 3.32. *A finite simple semigroup contains a single $\mathcal{D}$-class. This $\mathcal{D}$-class is regular and each of its $\mathcal{H}$-classes is a group.*

We can now state the main theorem of this section.

Theorem 3.33 (Rees-Sushkevich theorem)**.**

(1) *A finite semigroup is simple if and only if it is isomorphic to some Rees matrix semigroup.*

(2) *A finite semigroup is 0-simple if and only if it is isomorphic to some regular Rees matrix semigroup with zero.*

Proof. By Proposition 3.29, every regular Rees matrix semigroup with zero is 0-simple. Similarly by Proposition 3.30, every Rees matrix semigroup is simple.

Let S be a finite 0-simple semigroup. Let $(R_i)_{i\in I}$ $[(L_j)_{j\in J}]$ be the set of $\mathcal{R}$-classes [$\mathcal{L}$-classes] of S and let e be an idempotent of S. We let $H_{i,j}$ denote the $\mathcal{H}$-class $R_i \cap L_j$. By Propositions 1.13 and 3.31, the $\mathcal{H}$-class of e is a group G. Let us choose for each $i \in I$ an element $s_i \in L(e) \cap R_i$ and for each $j \in J$ an element $r_j \in R(e) \cap L_j$. By Proposition 1.3, $er_j = r_j$ and $s_i e = s_i$. Consequently, by Green's lemma the map $g \to s_i g r_j$ from G to $H_{i,j}$ is a bijection. It follows that each element of $S - 0$ admits a unique representation of the form $s_i g r_j$ with $i \in I$, $j \in J$ and $g \in G$.

Let $P = (p_{j,i})_{j\in J, i\in I}$ be the $J \times I$ matrix with entries in G^0 defined by $p_{j,i} = r_j s_i$. By Theorem 1.11, $r_j s_i \in G$ if $H_{i,j}$ contains an idempotent and $r_j s_i = 0$ otherwise. Define a map $\varphi : S \to M^0(G, I, J, P)$ by setting

$$\varphi(s) = \begin{cases} (i, g, j) & \text{if } s = s_i g r_j \\ 0 & \text{if } s = 0 \end{cases}$$

Clearly $\varphi(s)\varphi(0) = \varphi(0)\varphi(s) = 0 = \varphi(0)$. Let now s and s' be nonzero elements. Setting $\varphi(s) = (i, g, j)$ and $\varphi(s') = (i', g', j')$, we have

$$\varphi(s)\varphi(s') = (i, g, j)(i', g', j') = \begin{cases} (i, g r_j s_{i'} g', j) & \text{if } H_{i',j} \text{ contains an idempotent} \\ 0 & \text{otherwise} \end{cases}$$

Since $s \in H_{i,j}$ and $s' \in H_{i',j'}$, Theorem 1.11 shows that $ss' \neq 0$ if and only if $H_{i',j}$ contains an idempotent and in this case, $ss' = s_i g r_j s_{i'} g' r_j = s_i (g r_j s_{i'} g') r_j$. Therefore φ is a morphism, bijective by construction and hence is an isomorphism.

The case of finite simple semigroups can be handled in a similar way. □

Example 3.2. Let S be the semigroup generated by the following transformations:

	1	2	3	4	5	6
$* a$	1	1	1	4	4	4
$* b$	2	2	2	5	5	5
$* c$	3	3	3	6	6	6
$* d$	1	4	0	4	1	0

The elements of S and a set of relations defining S are given below

	1	2	3	4	5	6
$* a$	1	1	1	4	4	4
$* b$	2	2	2	5	5	5
$* c$	3	3	3	6	6	6
$* d$	1	4	0	4	1	0
bd	4	4	4	1	1	1
$* cd$	0	0	0	0	0	0
db	2	5	0	5	2	0
dc	3	6	0	6	3	0
bdb	5	5	5	2	2	2
bdc	6	6	6	3	3	3
dbd	4	1	0	1	4	0
$* dbdb$	5	2	0	2	5	0
$dbdc$	6	3	0	3	6	0

Relations:

$$\begin{array}{lllll} aa = a & ab = b & ac = c & ad = a & ba = a \\ bb = b & bc = c & ca = a & cb = b & cc = c \\ da = d & dd = d & cd = 0 & dcd = 0 & bdbd = a \end{array}$$

Finally, its $\mathcal{D}$-class structure is the following:

$^* a \quad bd$	$^* b \quad bdb$	$^* c \quad bdc$
$^* d \quad dbd$	$^* dbdb \quad db$	$dc \quad dbdc$

$$\boxed{^* 0}$$

It follows that S is a 0-simple semigroup, isomorphic to the Rees matrix semigroup $M(\mathbb{Z}/2\mathbb{Z}, 2, 3, P)$, where $\mathbb{Z}/2\mathbb{Z}$ is the multiplicative group $\{1, -1\}$ and

$$P = \begin{pmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 0 \end{pmatrix}$$

The isomorphism is given by $a = (1, 1, 1)$, $b = (1, 1, 2)$, $c = (1, 1, 3)$ and $d = (2, 1, 1)$.

The Rees-Sushkevich theorem has some particular cases of interest. If G is trivial and $P_{i,j} = 1$ for all $i \in I$ and $j \in J$, then $M(I, J, G, P)$ is isomorphic to a *rectangular band* $B(I, J)$, which is the set $I \times J$ with the multiplication

$$(i, j)(k, \ell) = (i, \ell)$$

If $I = \{1, \ldots, n\}$ and $J = \{1, \ldots, m\}$, the notation $B(n, m)$ is also used.

Corollary 3.34. *Let S be a nonempty finite aperiodic semigroup. The following conditions are equivalent:*

(1) *S is simple,*
(2) *S is idempotent and for all $e, f, s \in S$, $esf = ef$,*
(3) *S is isomorphic to a rectangular band.*

Proof. The equivalence of (1) and (3) follows directly from the Rees-Sushkevich theorem.

(3) implies (2) since in $B(I, J)$, $(i, j)(k, \ell)(\ell, m) = (i, m) = (i, j)(\ell, m)$.

(2) implies (1). Let I be a nonempty ideal of S and let $s \in I$ and $e \in S$. Since I is an ideal, $ese \in I$. Furthermore, one has by (2), $ese = ee = e$. It follows that $e \in I$ and hence $I = S$. Therefore S is simple. □

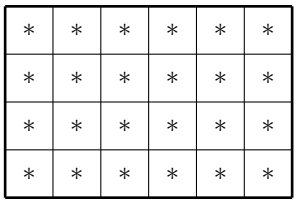


Figure 3.2. An aperiodic simple semigroup: the rectangular band $B(4, 6)$.

Furthermore, if I [J] is a singleton and then $M(I, J, G, P)$ is a right [left] zero semigroup. Conversely, any right [left] zero semigroup is isomorphic to such a Rees matrix semigroup.

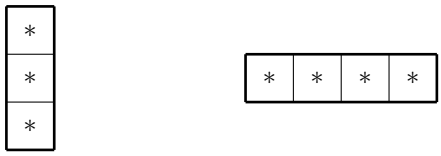


Figure 3.3. A left zero semigroup and a right zero semigroup.

4 Structure of regular $\mathcal{D}$-classes

Let D be a regular $\mathcal{D}$-class of a semigroup S. We define a semigroup D^0 whose support is $D \cup 0$ and multiplication (denoted by $*$) is given by

$$s * t = \begin{cases} st & \text{if } st \in D, \\ 0 & \text{otherwise} \end{cases}$$

We then have the following proposition.

Proposition 4.35. *If D is a regular $\mathcal{D}$-class of a semigroup, D^0 is a regular 0-simple semigroup.*

Proof. We first verify that all elements of D are $\mathcal{D}$-equivalent in D^0. Let $s, t \in D$ and let $r \in D$ be such that $s \mathcal{R} r \mathcal{L} t$. Let u and v be elements of S^1 such that $r = su$ and $s = rv$. Since D is regular, $L(s)$ $[L(r)]$ contains an idempotent e $[f]$. Thus $se = s$ and $rf = r$ by Proposition 1.3. It follows that $r = s(eu)$ and $s = r(fv)$. Furthermore, $eu \mathcal{L} su$ since $e \mathcal{L} s$ and thus $eu \in D$. Similarly, $fv \in D$ and hence $s \mathcal{R} r$ in D^0. Dually, $r \mathcal{L} t$ in D^0 and finally, $s \mathcal{D} t$ in D^0.

It follows that 0 and D^0 are the only ideals of D^0. Thus D^0 is 0-simple. Since D is regular, D^0 is also regular. □

An important particular case occurs when D is itself a semigroup. We say in this case that is a *full* $\mathcal{D}$-class of S.

Proposition 4.36. *Let D be a regular $\mathcal{D}$-class of a finite semigroup. The following conditions are equivalent:*

(1) *D is full,*
(2) *D is a simple semigroup,*
(3) *every $\mathcal{H}$-class of D contains an idempotent,*
(4) *the product of any two idempotents of D is also in D,*
(5) *D is isomorphic to some Rees matrix semigroup.*

Proof. The equivalence of (1) and (2) is trivial and that of (2) and (5) follows from Theorem 3.33.

(1) implies (4) is trivial.

(4) implies (3). Let H be an $\mathcal{H}$-class of D and let $x \in H$. Since D is regular, there exist by Proposition 1.12 an idempotent e in $R(x)$ and an idemptotent f in $L(x)$. By (4), fe belongs to D and thus, by the Location Theorem (Theorem 1.11), $R(e) \cap L(f)$ contains an idempotent. But $R(e) \cap L(f)$ is equal to H, which proves (3).

(3) implies (1) also follows from the Location Theorem. □

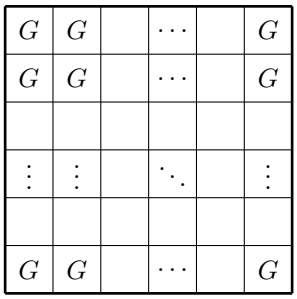


Figure 4.1. A completely regular $\mathcal{D}$-class.

4.1 Structure of the minimal ideal

Proposition 4.37. *Let S be a finite semigroup. Then S has a unique minimal ideal. This ideal is a regular simple semigroup.*

Proof. The set of all ideals has a $\leqslant_{\mathcal{J}}$-minimal element I, which is the unique minimal ideal of S. By construction, I is simple. Let $s \in S$. The descending sequence $s \geqslant_{\mathcal{J}} s^2 \geqslant_{\mathcal{J}} s^3 \dots$ is stationary. In particular, there exists an integer n such that $s^n \ \mathcal{J} \ s^{2n}$ and hence $s^n \ \mathcal{H} \ s^{2n}$. It follows by Proposition 1.13 that $H(s^n)$ contains an idempotent. Thus $E(S)$ is nonempty and contains a $\leqslant_{\mathcal{J}}$-minimal element e. This minimal idempotent belongs to I and thus I is regular. □

It follows that the minimal ideal of a finite semigroup has the following structure. Every $\mathcal{H}$-class is a group and all these groups are isomorphic.

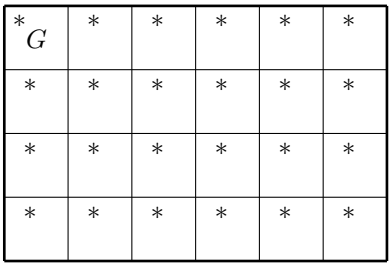


Figure 4.2. The minimal ideal of a finite semigroup.

5 Green's relations in subsemigroups and quotients

Let us start with a trivial observation: Green's relations are stable under morphisms.

Proposition 5.38. *Let $\varphi : S \to T$ be a morphism and let $\mathcal{K}$ be one of the relations $\leqslant_{\mathcal{R}}$, $\leqslant_{\mathcal{L}}$, $\leqslant_{\mathcal{H}}$, $\leqslant_{\mathcal{J}}$, $\mathcal{R}$, $\mathcal{L}$, $\mathcal{H}$, $\mathcal{D}$ or $\mathcal{J}$. If $s \ \mathcal{K} \ t$, then $\varphi(s) \ \mathcal{K} \ \varphi(t)$.*

Let now T be a subsemigroup [a quotient] of a semigroup S. It is often useful to compare Green's relations defined in S and T. For this purpose, if $\mathcal{K}$ is any one of Green's relations or preorders, we let $\mathcal{K}_S$ [$\mathcal{K}_T$] denote the Green's relation or preorder defined in the semigroup S [T].

5.1 Green's relations in subsemigroups

We first consider the case of subsemigroups.

Proposition 5.39. *Let T be a subsemigroup of a finite semigroup S and let $s, t \in T$ with t regular in T. Let $\mathcal{K}$ be one of the relations $\leqslant_{\mathcal{R}}$, $\leqslant_{\mathcal{L}}$, $\leqslant_{\mathcal{H}}$, $\mathcal{R}$, $\mathcal{L}$ or $\mathcal{H}$. If $s\ \mathcal{K}_S\ t$, then $s\ \mathcal{K}_T\ t$.*

Proof. Suppose that $s \leqslant_{\mathcal{R}_S} t$. If $\bar{t}$ is an inverse of t in T, then $t\ \mathcal{R}_T\ t\bar{t}$ and thus $s \leqslant_{\mathcal{R}_S} t\bar{t}$. Since $t\bar{t}$ is idempotent, it follows from Proposition 1.3 that $t\bar{t}s = s$. Thus $s \leqslant_{\mathcal{R}_T} t\bar{t}$ and finally $s \leqslant_{\mathcal{R}_T} t$. The proof for the other relations is similar. □

Proposition 5.39 does not extend to $\leqslant_{\mathcal{J}}$, $\mathcal{D}$ nor $\mathcal{J}$. Let S be, unsurprinsingly, the universal counterexample $B_2 = \{a, b, ab, ba, 0\}$, with $aba = a$, $bab = b$ and $a^2 = b^2 = 0$. Let $T = E(S) = \{ab, ba, 0\}$. Then $ab\ \mathcal{J}_S\ ba$, but $ab \not\leqslant_{\mathcal{J}_T} ba$.

However, if T is an ideal of S, the following property holds:

Proposition 5.40. *Let T be an ideal of a finite semigroup S and let $s, t \in T$ with s or t regular in T. Let $\mathcal{K}$ be one of the relations $\leqslant_{\mathcal{R}}$, $\leqslant_{\mathcal{L}}$, $\leqslant_{\mathcal{H}}$, $\leqslant_{\mathcal{J}}$, $\mathcal{R}$, $\mathcal{L}$, $\mathcal{H}$ or $\mathcal{J}$. If $s\ \mathcal{K}_S\ t$, then $s\ \mathcal{K}_T\ t$.*

Proof. Suppose that $s \leqslant_{\mathcal{J}_S} t$. Then $s = utv$ for some $u, v \in S^1$. If s is regular, let $\bar{s}$ be an inverse of s in T. Then $s = s\bar{s}s\bar{s}s = s\bar{s}utv\bar{s}s$. Since T is an ideal, $s\bar{s}u$ and $v\bar{s}s$ are elements of T. Thus $s \leqslant_{\mathcal{J}_T} t$. If t is regular, let $\bar{t}$ be an inverse of t in T. Then $s = utv = ut\bar{t}t\bar{t}tv$. Since T is an ideal, $ut\bar{t}$ and $\bar{t}tv$ are elements of T. Thus $s \leqslant_{\mathcal{J}_T} t$. The proof for the other relations is similar. □

If T is a local subsemigroup of S (see Exercise II.10), a similar result holds without any regularity assumption.

Proposition 5.41. *Let e be an idempotent of a finite semigroup S and let $T = eSe$. Let $\mathcal{K}$ be one of the relations $\leqslant_{\mathcal{R}}$, $\leqslant_{\mathcal{L}}$, $\leqslant_{\mathcal{H}}$, $\leqslant_{\mathcal{J}}$, $\mathcal{R}$, $\mathcal{L}$, $\mathcal{H}$ or $\mathcal{J}$. If two elements s and t of T satisfy $s\ \mathcal{K}_S\ t$, then $s\ \mathcal{K}_T\ t$.*

Proof. Suppose that $s \leqslant_{\mathcal{R}_S} t$. Then $s = tu$ for some $u \in S^1$. Since $s = ese$ and $t = ete$, $s = ese = etue = eeteue = eteeue = teue$. Thus $s \leqslant_{\mathcal{R}_T} t$. The proof for the other relations is similar. □

A useful consequence of Proposition 5.39 is the following corollary:

Corollary 5.42. *Let T be a subsemigroup of a finite semigroup S and let D be a regular $\mathcal{D}_T$-class of T. Then the restrictions to D of the Green's relations in S and T coincide.*

Proof. Since D is a $\mathcal{D}_T$-class of T, the relations $\mathcal{D}_T$, $\mathcal{D}_S$, $\mathcal{J}_T$ and $\mathcal{J}_S$ are universal on D and hence equal. The rest of the corollary follows directly from Proposition 5.39. □

5.2 Green's relations in quotient semigroups

In this subsection, φ will denote a surjective morphism from a semigroup S onto a semigroup T. Little can be said in the general case.

Proposition 5.43. *Let $\mathcal{K}$ be one of the relations $\mathcal{R}$, $\mathcal{L}$, $\mathcal{H}$, $\mathcal{D}$ or $\mathcal{J}$ and let K be a $\mathcal{K}_T$-class of T. Then $\varphi^{-1}(K)$ is a union of $\mathcal{K}_S$-classes.*

Proof. The result follows immediately from Proposition 5.38. □

Stronger results hold for finite semigroups.

Proposition 5.44. *Suppose that S is finite. Let J be a $\mathcal{J}$-class of T and let I be a $\leqslant_{\mathcal{J}}$-minimal $\mathcal{J}$-class of S contained in $\varphi^{-1}(J)$. Then*

(1) *$\varphi(I) = J$ and φ induces a surjective morphism from I^0 onto J^0,*
(2) *each $\mathcal{R}$-class [$\mathcal{L}$-class] of S contained in I maps under φ onto an $\mathcal{R}$-class [$\mathcal{L}$-class] of T contained in J,*
(3) *I is regular if and only if J is regular. In this case, I is the unique minimal $\mathcal{J}$-class of $\varphi^{-1}(J)$.*
(4) *If J is null, then every $\mathcal{J}$-class in $\varphi^{-1}(J)$ is null.*

Proof. (1) First, by Proposition 5.43, $\varphi^{-1}(J)$ is a union of $\mathcal{J}$-classes of S. Since S is finite, there exists a $\leqslant_{\mathcal{J}}$-minimal $\mathcal{J}$-class I of S contained in $\varphi^{-1}(J)$.

Let $s \in I$ and $t \in J$. Since $\varphi(s)\ \mathcal{J}\ t$ and φ is surjective, one has $t = \varphi(u)\varphi(s)\varphi(v)$ for some $u, v \in S^1$. Therefore $\varphi(usv) = t$ and $usv \in \varphi^{-1}(J)$. Since $s \in I$ and I is a minimal $\mathcal{J}$-class in $\varphi^{-1}(J)$, it follows that $usv \in I$. Therefore $\varphi(I) = J$.

Let $s, t \in I$. If $st \in I$, then $\varphi(st) \in J$. If $st \notin I$, then $st <_{\mathcal{J}} s$ and hence $\varphi(st) \notin J$: otherwise, I would not be a minimal $\mathcal{J}$-class in $\varphi^{-1}(J)$. Therefore, φ induces a well-defined surjective morphism from I^0 onto J^0.

(2) Let R be an $\mathcal{R}$-class of I. Let $r \in R$ and let t be an element of J $\mathcal{R}$-equivalent to $\varphi(r)$. Then $t = \varphi(r)x$ for some $x \in T^1$ and since φ is surjective, $\varphi(u) = x$ for some $u \in S^1$. It follows that $\varphi(ru) = \varphi(r)\varphi(u) = \varphi(r)x = t$ and therefore $ru \in \varphi^{-1}(J)$. Since $ru \leqslant_{\mathcal{R}} r$ and I is a minimal $\mathcal{J}$-class of $\varphi^{-1}(J)$ containing ru, one gets $ru \in I$. It follows by Theorem 1.9 that $ru \in R$ and thus $t \in \varphi(R)$. Consequently, φ maps R onto an $\mathcal{R}$-class of T contained in J. The proof is dual for the $\mathcal{L}$-classes.

(3) If I is regular, it contains an idempotent e. Therefore $\varphi(e)$ is an idempotent of J, and J is regular. Suppose now that J is regular an let e be an idempotent of J. By (1), there is an element $x \in I$ such that $\varphi(x) = e$. It follows that $\varphi(x^\omega) = e$ and since $x^\omega \leqslant x$ and I is a $\leqslant_{\mathcal{J}}$-minimal $\mathcal{J}$-class, one gets $x^\omega \in I$. Thus I is regular.

Let I_1 and I_2 be two $\leqslant_{\mathcal{J}}$-minimal $\mathcal{J}$-classes of $\varphi^{-1}(J)$. The previous argument shows that there exist two idempotents $e_1 \in I_1$ and $e_2 \in I_2$ such that $\varphi(e_1) = \varphi(e_2) = e$. It follows that $\varphi(e_1e_2) = e$ and thus $e_1e_2 \in \varphi^{-1}(J)$. But since $e_1e_2 \leqslant_{\mathcal{J}} e_1$ and $e_1e_2 \leqslant_{\mathcal{J}} e_2$, one gets $e_1e_2 \in I_1 \cap I_2$, a contradiction.

(4) is a consequence of (3) since a $\mathcal{J}$-class is null if and only if it is not regular. □

One says that an element y of T is *lifted* to an element x of S if $\varphi(x) = y$. Similarly, a subset Y of T is lifted to a subset X of S if $\varphi(X) = Y$.

Theorem 5.45 (Lifting lemma)**.** *If φ is a surjective morphism from a finite semigroup onto a finite semigroup, then*

(1) *every idempotent lifts to an idempotent,*
(2) *every $\mathcal{J}$-class lifts to a $\mathcal{J}$-class,*
(3) *every regular $\mathcal{R}$-class [$\mathcal{L}$-class, $\mathcal{H}$-class] lifts to a regular $\mathcal{R}$-class [$\mathcal{L}$-class, $\mathcal{H}$-class],*
(4) *every group lifts to a group.*

Proof. (1) has been proved several times, but we repeat the argument. If e is an idempotent of T, and x is an element of S such that $\varphi(x) = e$, then $\varphi(x^\omega) = e$.

(2) follows from Proposition 5.44.

(3) Let R be a regular $\mathcal{R}$-class of T and let $r \in R$. Let J be the regular $\mathcal{J}$-class of T containing R. By Proposition 5.44, there is a unique $\leqslant_{\mathcal{J}}$-minimal $\mathcal{J}$-class I of S contained in $\varphi^{-1}(J)$. Choose $s \in I$ such that $\varphi(s) = r$. By Proposition 5.44 again, the $\mathcal{R}$-class of s maps onto an $\mathcal{R}$-class of J, which is necessarily R. Similarly, the $\mathcal{L}$-class of s maps onto the $\mathcal{L}$-class of r and hence, the $\mathcal{H}$-class of s maps onto the $\mathcal{H}$-class of r.

(4) If H is a group of T, it is contained into a maximal group G of T which is also an $\mathcal{H}$-class. Then we may use the argument of (3) with r idempotent and we may choose by (1) s to be also idempotent. Then G lifts to an $\mathcal{H}$-class containing an idempotent, which is therefore a group K. In particular, φ induces a group morphism from K onto G and the group $\varphi^{-1}(H)$ is mapped onto H. This proves that H lifts to a group. □

The following example shows that Property (2) in Proposition 5.44 does not extend to $\mathcal{H}$-classes.

Example 5.1. The minimal automata of the languages

$$K = \{a^i b a^j \mid i \equiv 0 \bmod 2 \text{ and } j \equiv 0 \bmod 2\}$$
$$L = \{a^i b a^j \mid i + j \equiv 0 \bmod 2\}$$

are represented in Figure 5.1:

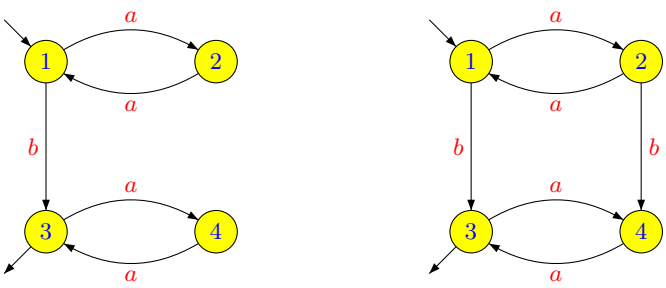


Figure 5.1. The minimal automata of K and L.

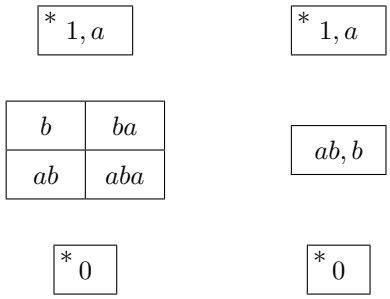
Their syntactic monoids M (on the left) and N (on the right) are respectively

1	1	2	3	4
a	2	1	4	3
b	3	0	0	0
ab	0	3	0	0
ba	4	0	0	0
bb	0	0	0	0
aba	0	4	0	0

1	1	2	3	4
a	2	1	4	3
b	3	4	0	0
ab	4	3	0	0
bb	0	0	0	0

The monoid $M = \{1, a, b, ab, ba, bab, 0\}$ is presented on $\{a, b\}$ by the relations $aa = 1$, $bab = 0$ and $bb = 0$. The monoid $N = \{1, a, b, ab, 0\}$ is presented on $\{a, b\}$ by the relations $aa = 1$, $ba = ab$ and $bb = 0$.

The $\mathcal{J}$-class structure of M and N is represented below:



Let $\varphi : M \to N$ be the surjective morphism defined by $\varphi(1) = 1$, $\varphi(a) = a$, $\varphi(b) = b$, $\varphi(ab) = ab$, $\varphi(ba) = ab$ and $\varphi(aba) = b$. Then $J = \{ab, b\}$ is both a $\mathcal{J}$-class and an $\mathcal{H}$-class of N and $I = \{b, ab, ba, aba\} = \varphi^{-1}(J)$ is a $\mathcal{J}$-class of M. However no $\mathcal{H}$-class of I is mapped onto J.

6 Green's relations and transformations

Semigroups are often given as transformation semigroups. We shall therefore examine the Green's relations in such semigroups.

Given an element $a \in \mathcal{T}(E)$, we denote by $\mathrm{Im}(a)$ the range of a and by $\mathrm{Ker}(a)$ the partition on E induced by the equivalence relation $\sim_a$ defined by

$$p \sim_a q \iff p \cdot a = q \cdot a$$

Finally, we set $\mathrm{rank}(a) = |\mathrm{Im}(a)| = |\mathrm{Ker}(a)|$. For example, if

$$a = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 1 & 4 & 5 & 5 & 5 & 4 & 1 \end{pmatrix}$$

we have $\mathrm{Im}(a) = \{1, 4, 5\}$, $\mathrm{Ker}(a) = 17/26/345$ and $\mathrm{rank}(a) = 3$.

We first mention an elementary property of the rank.

Lemma 6.46. *Let $a, b \in \mathcal{T}(E)$. Then* $\mathrm{rank}(ab) \leqslant \max\{\mathrm{rank}(a), \mathrm{rank}(b)\}$.

Proof. This follows from the two relations $\mathrm{Im}(ab) \subseteq \mathrm{Im}(b)$ and $\mathrm{Ker}(ab) \subseteq \mathrm{Ker}(a)$. □

One can now describe the Green's relations in the semigroup $\mathcal{T}(E)$.

Proposition 6.47. *Let a, b be elements of $\mathcal{T}(E)$. Then*

(1) *$a \leqslant_{\mathcal{R}} b$ if and only if $\mathrm{Ker}(a)$ is a partition coarser than $\mathrm{Ker}(b)$ and $a \mathcal{R} b$ if and only if $\mathrm{Ker}(a) = \mathrm{Ker}(b)$,*

(2) *$a \leqslant_{\mathcal{L}} b$ if and only if $\mathrm{Im}(a) \subseteq \mathrm{Im}(b)$ and $a \mathcal{L} b$ if and only if $\mathrm{Im}(a) = \mathrm{Im}(b)$,*

(3) *$a \leqslant_{\mathcal{J}} b$ if and only if $\mathrm{rank}(a) \leqslant \mathrm{rank}(b)$ and $a \mathcal{J} b$ if and only if $\mathrm{rank}(a) = \mathrm{rank}(b)$.*

Proof. (1) If $a \leqslant_{\mathcal{R}} b$, there exists $u \in \mathcal{T}(E)$, such that $a = bu$ and therefore $\mathrm{Ker}(a)$ is coarser than $\mathrm{Ker}(b)$. Conversely, if this condition is satisfied, the relation $u = a \circ b^{-1}$ is a function such that $bu = a$. Therefore $a \leqslant_{\mathcal{R}} b$. The result for $\mathcal{R}$ follows immediately.

(2) If $a \leqslant_{\mathcal{L}} b$, there exists $u \in \mathcal{T}(E)$, such that $a = ub$ and therefore $\mathrm{Im}(a) \subseteq \mathrm{Im}(b)$. Conversely, if $\mathrm{Im}(a) \subseteq \mathrm{Im}(b)$, there exists for each $q \in E$ an element q' such that $q' \cdot b = q \cdot a$. The function $q \to q'$ defines a transformation u such that $ub = a$ and thus $a \leqslant_{\mathcal{L}} b$. The result for $\mathcal{L}$ follows immediately.

(3) If $a \leqslant_{\mathcal{J}} b$, there exist $u, v \in \mathcal{T}(E)$ such that $a = ubv$ and therefore $\mathrm{rank}(a) \leqslant \mathrm{rank}(b)$. Conversely, suppose that $\mathrm{rank}(a) \leqslant \mathrm{rank}(b)$. We construct a transformation u by sending each class of $\mathrm{Ker}(a)$ onto an element of $\mathrm{Im}(b)$ and two distinct classes onto two distinct elements; this is possible since $|\mathrm{Im}(a)| = |\mathrm{Ker}(a)| \leqslant |\mathrm{Im}(b)|$. Then $\mathrm{Ker}(u) = \mathrm{Ker}(a)$ and $\mathrm{Im}(u) \subseteq \mathrm{Im}(b)$ by construction. Therefore $a \mathcal{R} u$ by (1), $u \leqslant_{\mathcal{L}} b$ by (2) and finally $a \leqslant_{\mathcal{J}} u \leqslant_{\mathcal{J}} b$. The result for $\mathcal{J}$ follows immediately. □

We can now pass to the general study of transformation semigroups. Given a set E and a partition $\mathcal{E} = \{E_1, \ldots, E_n\}$ of E, we say that a subset F of E is a *transversal* of $\mathcal{E}$ if, for $1 \leqslant i \leqslant n$, $|E_i \cap F| = 1$.

Proposition 6.48. *Let S be a subsemigroup of $\mathcal{T}(E)$. An element a belongs to a group of S if and only if $\mathrm{Im}(a)$ is a transversal of $\mathrm{Ker}(a)$.*

Proof. If a belongs to a group, then $a^n = a$ for a certain $n \geqslant 2$ and therefore a induces a bijection on $\mathrm{Im}(a)$. Let K be a class of $\mathrm{Ker}(a)$. If $|K \cap \mathrm{Im}(a)| \geqslant 2$, two elements of $\mathrm{Im}(a)$ have the same image under a, which contradicts the above property. Therefore $|K \cap \mathrm{Im}(a)| \leqslant 1$ for every $K \in \mathrm{Ker}(a)$. Furthermore, if $K \cap \mathrm{Im}(a) = \emptyset$ for a class K of $\mathrm{Ker}(a)$, it follows that

$$|\mathrm{Im}(a)| = \sum_{K \in \mathrm{Ker}(a)} |K \cap \mathrm{Im}(a)| < |\mathrm{Ker}(a)|$$

a contradiction. Therefore $\mathrm{Im}(a)$ is a transversal of $\mathrm{Ker}(a)$.

Conversely, if this condition is satisfied, a induces a bijection on its image and therefore $a^n = a$ for some $n \geqslant 2$. It follows by Proposition 1.13 that a belongs to a group of S. □

Part of Proposition 6.47 extends to all transformation semigroups.

Proposition 6.49. *Let S be a subsemigroup of $\mathcal{T}(E)$. Then*

(1) *if $a \leqslant_{\mathcal{R}} b$, then $\mathrm{Ker}(a)$ is a coarser partition than $\mathrm{Ker}(b)$ and if $a \mathcal{R} b$, then $\mathrm{Ker}(a) = \mathrm{Ker}(b)$,*

(2) *if $a \leqslant_{\mathcal{L}} b$, then $\mathrm{Im}(a) \subseteq \mathrm{Im}(b)$ and if $a \mathcal{L} b$, then $\mathrm{Im}(a) = \mathrm{Im}(b)$,*

(3) *if* $a \leqslant_{\mathcal{J}} b$, *then* $\operatorname{rank}(a) \leqslant \operatorname{rank}(b)$ *and if* $a \mathcal{J} b$, *then* $\operatorname{rank}(a) = \operatorname{rank}(b)$.

Proof. See the proof of Proposition 6.47. □

Note that the equivalences stated in Proposition 6.47 for $\mathcal{T}(E)$ **do not hold** for all transformation semigroups. For instance, in the semigroup S represented in Figure A.2, the elements acb and $acbc$ have the same image but are not $\mathcal{L}$-equivalent. Similarly, $\operatorname{rank}(acac) = \operatorname{rank}(cbcb)$, but these two idempotents are not $\mathcal{J}$-equivalent.

However, Proposition 6.48 enables us to locate very easily the elements of a group in S. One can then completely reconstruct the regular $\mathcal{D}$-class of such an element x by the following algorithm:

(1) Calculate all the sets of the form $\operatorname{Im}(xr)$ (where $r \in S^1$) such that $|\operatorname{Im}(xr)| = |\operatorname{Im}(x)|$. For each such set I, remember an element r such that $I = \operatorname{Im}(xr)$.

(2) In parallel to (1), calculate all the transformations xr ($r \in S^1$) such that $\operatorname{Im}(xr) = \operatorname{Im}(x)$. One obtains the $\mathcal{H}$-class of x.

(3) Calculate all the partitions of the form $\operatorname{Ker}(sx)$ (where $s \in S^1$) such that $|\operatorname{Ker}(sx)| = |\operatorname{Ker}(x)|$. For each such partition K, remember an element s such that $K = \operatorname{Ker}(sx)$.

(4) Among the sets calculated in (1), retain only those which are transversals of one of the partitions calculated in (3): one obtains a set $\mathcal{I} = \{\operatorname{Im}(xr_1), \ldots, \operatorname{Im}(xr_n)\}$, where $r_1 = 1$. Similarly, among the partitions calculated in (3), retain only those which admit as a transversal one of the sets calculated in (1). We obtain a set $\mathcal{K} = \{\operatorname{Ker}(s_1x), \ldots, \operatorname{Ker}(s_mx)\}$, where $s_1 = 1$.

(5) The results are summarised in the double-entry table below, representing the egg-box picture of the $\mathcal{D}$-class of x. If $H_{i,j} = R_{s_ix} \cap L_{xr_j}$, one has $H_{i,j} = s_iHr_j$, which enables us to calculate the $\mathcal{D}$-class completely. Finally, the $\mathcal{H}$-class $H_{i,j}$ is a group if and only if $\operatorname{Im}(xr_j)$ is a transversal of $\operatorname{Ker}(s_ix)$.

	$\operatorname{Im}(xr_1)$		$\operatorname{Im}(xr_j)$		$\operatorname{Im}(xr_n)$
	H		xr_j		
$\operatorname{Ker}(x) = \operatorname{Ker}(s_1x)$	s_ix		$H_{i,j}$		
$\operatorname{Ker}(s_mx)$	s_mx				

Justification of the algorithm.

Recall that x is an element of a group in S. Suppose that $rx \mathcal{R} x$ and let $e \in E(S)$ and $s \in S^1$ are such that $x \mathcal{R} xr$, $xr \mathcal{L} e$, $e \mathcal{R} sx$ and $sx \mathcal{L} x$.

x		xr
sx		e

then by Proposition 6.48 $\mathrm{Im}(e)$ is a transversal of $\mathrm{Ker}(e)$ and by Proposition 6.49, $\mathrm{Im}(e) = \mathrm{Im}(xr)$ and $\mathrm{Ker}(sx) = \mathrm{Ker}(e)$, so that $\mathrm{Im}(xr)$ is a transversal of $\mathrm{Ker}(sx)$. Converserly, suppose that $\mathrm{Im}(xr)$ is a transversal of a certain $\mathrm{Ker}(sx)$, where $|\mathrm{Im}(xr)| = |\mathrm{Im}(x)|$ and $|\mathrm{Ker}(sx)| = |\mathrm{Ker}(x)|$. Let $e \in \mathcal{T}(E)$ be a transformation with range $\mathrm{Im}(xr)$ and with kernel $\mathrm{Ker}(sx)$. Then by Proposition 6.47 one has in $\mathcal{T}(E)$, $x \mathcal{D} xr \mathcal{D} sx$, $xr \mathcal{L} e$ and $e \mathcal{L} sx$. It follows that $x \mathcal{R} xr$, $xr \mathcal{L} e$, $e \mathcal{R} sx$ and $sx \mathcal{L} x$. Now, x, xr and sx belong to S and by Proposition 2.27, one has $e \in S$ and x, xr, sx and e are in the same $\mathcal{D}$-class of S.

This proves that the algorithm computes the $\mathcal{R}$-class of x.

In practice, this algorithm is only useful for computing regular $\mathcal{D}$-classes.

7 Summary: a complete example

We compute in this section the minimal ordered automaton and the syntactic ordered monoid of the language $L = (a + bab)^*(1 + bb)$. We invite the reader to verify this computation by hand. It is the best possible exercise to master the notions introduced in the first chapters.

The minimal automaton of L is represented in Figure 7.1.

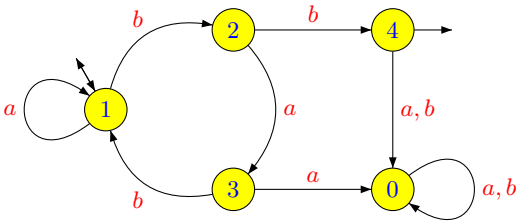


Figure 7.1. The minimal automaton of L.

The order on the set of states is $2 < 4$ and $0 < q$ for each state $q \neq 0$. The

syntactic monoid of L has 27 elements including a zero: $b^2a^2 = 0$.

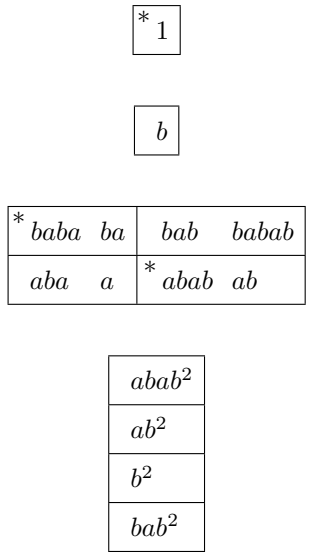
	1	2	3	4
$* 1$	1	2	3	4
a	1	3	0	0
b	2	4	1	0
$* a^2$	1	0	0	0
ab	2	1	0	0
ba	3	0	1	0
b^2	4	0	2	0
a^2b	2	0	0	0
aba	3	1	0	0
ab^2	4	2	0	0
ba^2	0	0	1	0
bab	1	0	2	0
$* b^2a$	0	0	3	0
b^3	0	0	4	0

	1	2	3	4
a^2ba	3	0	0	0
a^2b^2	4	0	0	0
aba^2	0	1	0	0
$* abab$	1	2	0	0
ab^2a	0	3	0	0
ab^3	0	4	0	0
ba^2b	0	0	2	0
$* baba$	1	0	3	0
bab^2	2	0	4	0
$* b^2a^2$	0	0	0	0
$* aba^2b$	0	2	0	0
$abab^2$	2	4	0	0
$babab$	2	0	1	0

It is defined by the following set of relations:

$$a^2a = a^2 \qquad b^2ab = ba^2 \qquad b^3a = 0 \qquad b^4 = 0 \qquad a^2ba^2 = 0$$
$$a^2bab = a^2 \qquad a^2b^2a = 0 \qquad a^2b^3 = 0 \qquad ababa = a \qquad ab^2a^2 = 0$$
$$ba^2ba = b^2a \qquad ba^2b^2 = b^3 \qquad baba^2 = a^2 \qquad bab^2a = a^2ba \qquad bab^3 = a^2b^2$$
$$b^2a^2 = 0 \qquad babab^2 = b^2$$

It contains 7 $\mathcal{D}$-classes, 15 $\mathcal{R}$-classes, 10 $\mathcal{L}$-classes and 23 $\mathcal{H}$-classes. The minimal ideal reduces to the zero, but there is a unique regular 0-minimal ideal. The $\mathcal{D}$-class structure is represented below:

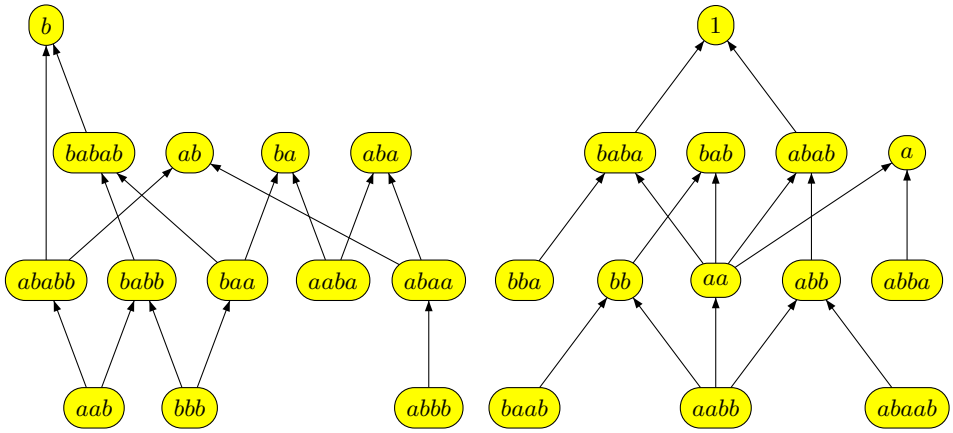


* a^2	a^2ba	a^2b
ba^2	* b^2a	ba^2b
aba^2	ab^2a	* aba^2b

a^2b^2
ab^3
b^3

* 0

The syntactic preorder satisfies $0 < x$ for all $x \neq 0$. The other relations are represented below, with our usual convention: an arrow from u to v means that $u < v$.



8 Exercises

Exercise 1. Describe Green's relations in the syntactic monoids of Exercise IV.8.

Exercise 2. Verify that in the example of Section 7, $a \mathcal{H} aba$ but $a^\omega \neq (aba)^\omega$.

Exercise 3. Describe Green's relations in the following monoids:

(1) the monoid of all functions from $\{1, 2, 3\}$ to itself,
(2) the monoid of all order-preserving functions from $\{1, 2, 3\}$ to itself (a function f is *order-preserving* if $i \leqslant j$ implies $f(i) \leqslant f(j)$),
(3) the monoid of all partial functions from $\{1, 2, 3\}$ to itself,
(4) the monoid of all relations on $\{1, 2, 3\}$,
(5) the monoid of all reflexive relations on $\{1, 2, 3\}$,
(6) the monoid of all upper-triangular Boolean matrices of size 3×3,

(7) the monoid of all unitriangular Boolean matrices of size 3×3 (a matrix is unitriangular if it is upper-triangular and if its diagonal entries are all equal to 1).

Exercise 4. Show that there exist finite semigroups in which none of Green's relations is a congruence.

Exercise 5. Let M be a finite monoid. Let $\mathcal{P}(M)$ be the set of subsets of M.

(1) Show that $\mathcal{P}(M)$ is a monoid for the following product: given two subsets X and Y, $XY = \{xy \mid x \in X,\ y \in Y\}$.

(2) Show that the set $\mathcal{P}_1(M)$ of subsets of M containing 1 is a submonoid of M. Show that this monoid is $\mathcal{J}$-trivial.

Exercise 6. Let s and t be regular elements of a semigroup S. Show that the following conditions are equivalent:

(1) $s \mathcal{R} t$,

(2) there exists $\bar{s} \in V(s)$ and $\bar{t} \in V(t)$ such that $s\bar{s} = t\bar{t}$,

(3) for all $\bar{s} \in V(s)$, there exists $\bar{t} \in V(t)$ such that $s\bar{s} = t\bar{t}$.

A dual result holds for $\mathcal{L}$. Finally, show that the following conditions are equivalent:

(1) $s \mathcal{H} t$,

(2) there exists $\bar{s} \in V(s)$ and $\bar{t} \in V(t)$ such that $s\bar{s} = t\bar{t}$ and $\bar{s}s = \bar{t}t$,

(3) for all $\bar{s} \in V(s)$, there exists $\bar{t} \in V(t)$ such that $s\bar{s} = t\bar{t}$ and $\bar{s}s = \bar{t}t$.

Exercise 7. Let a and b be two elements of a finite semigroup. Show that if $abb = b$, then $bab = b$.

Exercise 8. Let S be a nonempty finite semigroup. Suppose that, for all $a, b \in S$, the equation $ax = b$ has a solution and that the equation $xa = b$ also has a solution. Show that S is a group.

Exercise 9. Let S be a finite semigroup. Show that the number of inverses of an element s of S is equal to the product of the number of idempotents of the $\mathcal{R}$-class of s by the number of idempotents of the $\mathcal{L}$-class of s.

Exercise 10 (See [68])**.** Show that a monoid is a group if and only if every element has exactly one weak inverse. What happens for a semigroup?

Exercise 11. Let S be a regular semigroup. Show that the following conditions are equivalent:

(1) S is simple,

(2) for all $s \in S$, every weak inverse of s is also an inverse of s,

(3) for all s, t in S, if $us = ut$ and $sv = tv$ for some $u, v \in S$, then $s = t$.

Exercise 12. A *right group* is a semigroup that is right simple and left cancellative. Prove that if G is a group and S is a right zero semigroup, then $G \times S$ is a right group.

Show that the following conditions are equivalent:

(1) S is a right group,

(2) S is right simple and contains at least one idempotent,

(3) for each $s, t \in S$, the equation $ax = b$ has a unique solution in S,

(4) S is isomorphic to a Rees matrix semigroup $M(G, I, J, P)$, where $|I| = 1$ and P is the $J \times I$ column matrix whose entries are all equal to the identity of G.

Exercise 13. The $\mathcal{D}$-class structure of a finite semigroup is given below:

$^*a_1$	a_2	a_3	a_4	$^*a_5$
a_6	$^*a_7$	$^*a_8$	a_9	a_{10}
a_{11}	a_{12}	a_{13}	$^*a_{14}$	a_{15}

^{*0}

Compute $a_{11}a_2$, a_9a_{11}, a_7a_6, $a_{13}a_8$, a_7a_8, a_6a_7, a_8a_{13}, a_8a_7.

Exercise 14. A finite semigroup contains the following 0-minimal $\mathcal{D}$-class:

a	ac

^{*0}

Compute aa and aca. Show that there exists an idempotent e such that $ae = a$.

Exercise 15. A monoid M is an *inverse* monoid if every element of M has a unique inverse. Show that a finite monoid is inverse if and only if M is regular (that is, every element of M is regular) and the idempotents commute in M.

Exercise 16. Let Q be a finite set and let $\mathcal{I}(Q)$ be the monoid of partial injective functions from Q to Q under composition. Show that $\mathcal{I}(Q)$ is an inverse monoid and that if M is a finite inverse monoid, then M is isomorphic to a submonoid of $\mathcal{I}(M)$.

9 Notes

The material presented in this chapter can be found in almost every book on semigroups. My favourite references are the books of Lallement [78] and Grillet [55]. More information on semigroup theory can be found in Clifford and Preston [30, 31] and Higgins [60]. The book [136] by Rhodes and Steinberg contains much more advanced material and Green's relations are only treated in the appendix. We just treated the case of finite 0-simple semigroups in this chapter, but the structure of arbitrary completely 0-simple semigroups was elucidated by Rees.

Part B

Historical results

Chapter VI

Star-free languages

The characterisation of star-free languages, obtained by Schützenberger in 1965, is the second most important result of the theory of finite automata, right after Kleene's theorem.

1 Star-free languages

Let A be a finite alphabet. The set of *star-free* languages of A^* is the smallest set $\mathcal{R}$ of languages of A^* such that

(a) $\mathcal{R}$ contains the empty set, the set $\{1\}$ and, for each $a \in A$, the singleton $\{a\}$.

(b) $\mathcal{R}$ is closed under finite union, finite product and complement.

Thus the definition of the star-free languages follows the same definition scheme as the one of rational languages, with the difference that the star operation is replaced by the complement. Since the rational languages are closed under complement, every star-free language is rational, but we shall see later on that the converse is not true.

It follows immediately from the definition that every finite set is star-free. Indeed, every singleton $\{a_1 \cdots a_n\}$ can be written as the product $\{a_1\} \cdots \{a_n\}$ and every finite language is a finite union of singletons.

We shall follow the notation of Chapter IV. Union will be denoted additively, the empty set will be denoted by 0, the singleton $\{u\}$ will be simply denoted by u, the product will be denoted by simple juxtaposition and L^c will denote the complement of a subset L of A^*. The star-free sets are thus described by expressions using the letters of the alphabet A, the constants 0 and 1 and the three operators union, product and complement. It is not always easy to find such an expression, as is shown in the examples below.

Example 1.1.

(1) A^* is a set star-free, since $A^* = 0^c$

(2) If B is a subset of A, A^*BA^* is star-free by (1). It follows that B^* is star-free, since

$$B^* = A^* - \sum_{a \in A - B} A^* a A^* = \Big(\sum_{a \in A - B} 0^c a 0^c \Big)^c$$

(3) If $A = \{a, b\}$, the set $(ab)^*$ is star-free. Indeed

$$(ab)^* = (b0^c + 0^c a + 0^c aa0^c + 0^c bb0^c)^c$$

2 Aperiodic monoids

A finite monoid M is *aperiodic* if for each $x \in M$, there is an integer $n > 0$ such that $x^n = x^{n+1}$. Other characterisations of aperiodic monoids are given in Proposition 2.1 (see also Proposition V.2.26).

Proposition 2.1. *Let M be a finite monoid. The following conditions are equivalent:*

(1) *M is aperiodic,*
(2) *there is an integer $n > 0$ such that, for all $x \in M$, $x^n = x^{n+1}$,*
(3) *M is $\mathcal{H}$-trivial,*
(4) *the groups in M are trivial.*

Proof. (1) implies (2). By definition, for each $x \in M$, there is an integer $n_x > 0$ such that $x^{n_x} = x^{n_x+1}$. Let $n = \max_{x \in M} n_x$. Then $x^{n_x} x^{n-n_x} = x^{n_x+1} x^{n-n_x}$, that is, $x^n = x^{n+1}$, which proves (2).

(2) implies (1) is trivial.

Let now n be the exponent of M.

(2) implies (3). Suppose $a \mathcal{H} b$. Then there exist $u, v, x, y \in S^1$ such that $ua = b$, $vb = a$, $ax = b$ and $by = a$, whence $uay = a$ and therefore $u^n ay^n = a$. Since $u^n = u^{n+1}$, one gets $a = u^n ay^n = u^{n+1} ay^n = u(u^n ay^n) = ua = b$. Therefore M is $\mathcal{H}$-trivial.

(3) implies (4) follows from the fact that each group in M is contained in an $\mathcal{H}$-class.

(4) implies (2). Let $x \in M$. Then the $\mathcal{H}$-class of the idempotent x^n is a group G, which is trivial by (4). Since x^n and x^{n+1} belong to G, one gets $x^n = x^{n+1}$. □

Let us also prove a useful property of aperiodic monoids.

Proposition 2.2. *A finite ordered monoid M is aperiodic if and only if it satisfies the identity $x^{n+1} \leqslant x^n$ for some $n > 0$.*

Proof. If M is aperiodic, it satisfies by definition an identity of the form $x^{n+1} = x^n$ and the identity $x^{n+1} \leqslant x^n$ is trivially satisfied. Conversely, suppose that M satisfies the identity $x^{n+1} \leqslant x^n$ for some $n > 0$. Let ω be a multiple of the exponent of M such that $\omega \geqslant n$. Then

$$x^\omega = x^{2\omega} \leqslant x^{2\omega-1} \leqslant x^{2\omega-2} \leqslant \ldots \leqslant x^{\omega+1} \leqslant x^\omega$$

whence $x^\omega = x^{\omega+1}$ for all $x \in M$. Thus M is aperiodic. □

3 Schützenberger's theorem

We are now ready to state Schützenberger's theorem.

Theorem 3.3 (Schützenberger)**.** *A language is star-free if and only if its syntactic monoid is aperiodic.*

Proof. The easiest part of the proof relies on a syntactic property of the concatenation product[1]. Let L_0 and L_1 be two recognisable languages of A^* and let $L = L_0L_1$. Let M_0, M_1 and M be the syntactic ordered monoids of L_0, L_1 and L.

Lemma 3.4. *If M_0 and M_1 are aperiodic, so is M.*

Proof. Let n_0, n_1 and m be the respective exponents of M_0, M_1 and M and let $n = n_0 + n_1 + 1$. We claim that, for all $x \in M$, $x^{n+1} \leqslant x^n$. By Proposition 2.2, this property will suffice to show that M is aperiodic.

By the definition of the syntactic order, the claim is equivalent to proving that, for each $x, u, v \in A^*$, $ux^nv \in L$ implies $ux^{n+1}v \in L$. One can of course suppose that $x \neq 1$. If $ux^nv \in L$, there exists a factorisation $ux^nv = x_0x_1$ with $x_0 \in L_0$ and $x_1 \in L_1$. Two cases are possible. Either $x_0 = ux^{n_0}r$ with $rx_1 = x^{n-n_0}v$, or $x_1 = sx^{n_1}v$ with $x_0s = ux^{n-n_1}$. Let us consider the first case, since the second case is symmetric. Since M_0 is aperiodic and since $ux^{n_0}r \in L_0$, we have $ux^{n_0+1}r \in L_0$ and hence $ux^{n+1}v \in L$. □

Let us fix an alphabet A and let $\mathcal{A}(A^*)$ be the set of recognisable languages of A^* whose syntactic monoid is aperiodic. An elementary computation shows that the syntactic monoid of the languages $\{1\}$ and a, for $a \in A$, is aperiodic. Therefore, the set $\mathcal{A}(A^*)$ contains the languages of this type. Furthermore, by Proposition IV.2.9, a language and its complement have the same syntactic monoid, $\mathcal{A}(A^*)$ is closed under complement. It is also closed under finite union by Proposition IV.2.10 and hence under Boolean operations. Lemma 3.4 shows that $\mathcal{A}(A^*)$ is also closed under product. Consequently, $\mathcal{A}(A^*)$ contains the star-free sets.

To establish the converse[2], we need two elementary properties of aperiodic monoids. The first property is a simple reformulation of Theorem V.1.9 (5) in the case of aperiodic monoids.

Lemma 3.5 (Simplification Lemma)**.** *Let M an aperiodic monoid and let $p, q, r \in M$. If $pqr = q$, then $pq = q = qr$.*

Proof. Let n the exponent of M. Since $pqr = q$, we also have $p^nqr^n = q$. Since M is aperiodic, we have $p^n = p^{n+1}$ and hence $pq = pp^nqr^n = p^nqr^n = q$ and, in the same way, $qr = q$. □

The second property leads to a decomposition of each subset of an aperiodic monoid as a Boolean combination of right ideals, left ideals, or ideals.

Lemma 3.6. *Let M be an aperiodic monoid and let $m \in M$. Then $\{m\} = (mM \cap Mm) - J_m$, with $J_m = \{s \in M \mid m \notin MsM\}$.*

Proof. It is clear that $m \in (mM \cap Mm) - J_m$. Conversely, if $s \in (mM \cap Mm) - J_m$, there exist $p, r \in M$ such that $s = pm = mr$. Moreover, as $s \notin J_m$, $m \in MsM$. It follows by Theorem V.1.9 that $m \mathcal{H} s$ and by Proposition 2.1 that $m = s$ since M is aperiodic. □

[1] an improved version of this result is given in Theorem XVI.6.20

[2] Another proof is given on page 295.

We now need to prove that if $\varphi : A^* \to M$ is a morphism from A^* to an aperiodic monoid M, the set $\varphi^{-1}(P)$ is star-free for every subset P of M. The formula

$$\varphi^{-1}(P) = \sum_{m \in P} \varphi^{-1}(m)$$

allows us to assume that $P = \{m\}$. We shall show that $\varphi^{-1}(m)$ is star-free by induction on the integer $r(m) = |M - MmM|$. The initial step is treated in the next lemma.

Lemma 3.7. *If $r(m) = 0$, then $m = 1$ and $\varphi^{-1}(m)$ is star-free*

Proof. If $r(m) = 0$, then $M = MmM$ and there exist $u, v \in M$ such that $umv = 1$. Furthermore, the Simplification Lemma applied to $(um)1(v) = 1$ and to $(u)1(mv) = 1$ gives $u = v = 1$ and hence also $m = 1$. Let us show that $\varphi^{-1}(1) = B^*$, where $B = \{a \in A \mid \varphi(a) = 1\}$. If $u \in B^*$, we have of course $\varphi(u) = 1$. Conversely, if $\varphi(u) = 1$, the Simplification Lemma shows that $\varphi(a) = 1$ for each letter a of u, and hence $u \in B^*$. Now, as was shown in Example 1.1, (2), B^* is a star-free set. □

Assume now that $r(m) > 0$ and that the property has been established for each element s such that $r(s) < r(m)$. We shall now prove the formula

$$\varphi^{-1}(m) = (UA^* \cap A^*V) - (A^*CA^* \cup A^*WA^*) \tag{3.1}$$

where

$$U = \sum_{(n,a) \in E} \varphi^{-1}(n)a \qquad V = \sum_{(a,n) \in F} a\varphi^{-1}(n)$$
$$C = \{a \in A \mid m \notin M\varphi(a)M\} \qquad W = \sum_{(a,n,b) \in G} a\varphi^{-1}(n)b$$

with

$$E = \{(n, a) \in M \times A \mid n\varphi(a) \mathcal{R}\, m \text{ but } n \notin mM\}$$
$$F = \{(a, n) \in A \times M \mid \varphi(a)n \mathcal{L}\, m \text{ but } n \notin Mm\}$$
$$G = \{(a, n, b) \in A \times M \times A \mid$$
$$m \in (M\varphi(a)nM \cap Mn\varphi(b)M) - M\varphi(a)n\varphi(b)M\}$$

Denote by L the right-hand side of (3.1). We first prove the inclusion $\varphi^{-1}(m) \subseteq L$. Let $u \in \varphi^{-1}(m)$ and let p be the shortest prefix of u such that $\varphi(p) \mathcal{R}\, m$. The word p cannot be empty, since otherwise $m \mathcal{R}\, 1$, whence $m = 1$ by the Simplification Lemma. Put $p = ra$ with $r \in A^*$ and $a \in A$ and let $n = \varphi(r)$. By construction, $(n, a) \in E$ since

(a) $n\varphi(a) = \varphi(r)\varphi(a) = \varphi(p) \mathcal{R}\, m$,

(b) since $m \leqslant_{\mathcal{R}} \varphi(p) = n\varphi(a) \leqslant_{\mathcal{R}} n$, one has $n \notin mM$, for otherwise we would have $n \mathcal{R}\, m$.

It follows that $p \in \varphi^{-1}(n)a$ and $u \in UA^*$. A symmetric argument shows that $u \in A^*V$. If $u \in A^*CA^*$, there exists a letter a of C such that $m = \varphi(u) \in M\varphi(a)M$, a contradiction to the definition of C. Similarly, if $u \in A^*WA^*$, there exist $(a, n, b) \in G$ such that $m \in M\varphi(a)n\varphi(b)M$, a contradiction this time to the definition of G. Therefore $u \in L$.

Conversely, let $u \in L$ and let $s = \varphi(u)$. Since $u \in UA^*$, we have $u \in \varphi^{-1}(n)aA^*$ for some $(n, a) \in E$ and hence $s = \varphi(u) \in n\varphi(a)M$. Now, since $(n, a) \in E$, $n\varphi(a)M = mM$ and thus $s \in mM$. A dual argument shows that $u \in VA^*$ implies $s \in Mm$. By Lemma 3.6, in order to prove that $s = m$, and hence that $u \in \varphi^{-1}(m)$, it suffices to prove that $s \notin J_m$, that is, $m \in MsM$. Supposing the contrary, consider a factor f of u of minimal length such that $m \notin M\varphi(f)M$. The word f is necessarily nonempty. If f is a letter, this letter is in C and $u \in A^*CA^*$, which is impossible. We may thus set $f = agb$, with $a, b \in A$. Set $n = \varphi(g)$. Since f is of minimal length, we have $m \in M\varphi(a)nM$ and $m \in Mn\varphi(b)M$. Consequently $(a, n, b) \in G$ and $f \in W$, which is impossible again.

Formula (3.1) is thus established and it suffices now to show that U, V and W are star-free, since we have already seen in Example 1.1 that A^*CA^* is star-free. Let $(n, a) \in E$. Since $n\varphi(a)M = mM$, we have $MmM \subseteq MnM$ and hence $r(n) \leqslant r(m)$. Moreover, as $m \leqslant_{\mathcal{R}} n$, Theorem V.1.9 shows that if $MmM = MnM$, we have $n \mathcal{R} m$, which is not possible since $n \notin mM$. Therefore $r(n) < r(m)$ and U is star-free by the induction hypothesis.

A symmetric argument works for V. It remains to treat the case of W. Let $(a, n, b) \in G$. One has $r(n) \leqslant r(m)$ since $m \in MnM$. Suppose that $MmM = MnM$. Then in particular $n \in MmM$ and as $m \in M\varphi(a)nM$ and $m \in Mn\varphi(b)M$, it follows $n \in M\varphi(a)nM$ and $n \in Mn\varphi(b)M$, whence $n \mathcal{L} \varphi(a)n$ and $n \mathcal{R} n\varphi(b)$. By Proposition V.1.10, $n\varphi(b) \mathcal{L} \varphi(a)n\varphi(b)$, and hence $m \mathcal{J} \varphi(a)n\varphi(b)$, a contradiction to the definition of G. Consequently $r(n) < r(m)$ and W is star-free by the induction hypothesis. □

Example 3.1. Let $A = \{a, b\}$ and let $L = (ab)^*$. The minimal automaton of L is represented in Figure 3.1.

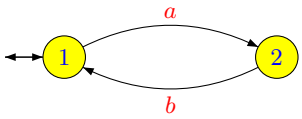


Figure 3.1. The minimal automaton of $(ab)^*$.

The syntactic monoid M of L is the monoid consisting of the six matrices

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad a = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \quad b = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$$
$$aa = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \quad ab = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad ba = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

and it is defined by the relations $a^2 = b^2 = 0$, $aba = a$ and $bab = b$. Its $\mathcal{D}$-class structure is given in Figure 3.2:

* 1

* ab	a
b	* ba

* 0

Figure 3.2. The $\mathcal{D}$-class structure of M.

This monoid is aperiodic, since $x^2 = x^3$ for each $x \in M$, and hence L is star-free.

Example 3.2. Let $A = \{a, b\}$ and let $L' = (aa)^*$. The minimal automaton of L' is represented in Figure 3.3:

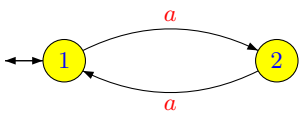


Figure 3.3. The minimal automaton of $(aa)^*$.

The syntactic monoid M' of L' is the monoid consisting of the three matrices

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad a = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad b = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

and it is defined by the relations $a^2 = 1$ and $b = 0$. Its $\mathcal{D}$-class structure is given in Figure 3.4:

* $1, a$

* 0

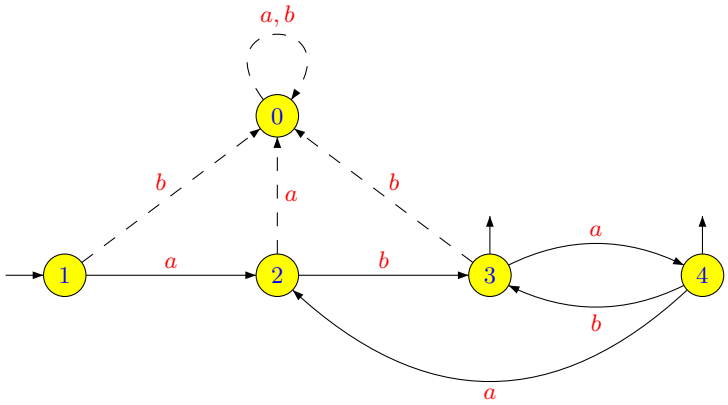
Figure 3.4. The $\mathcal{D}$-class structure of M'.

This monoid is not aperiodic, since, for each $n > 0$, $a^n \neq a^{n+1}$ and hence L' is not star-free.

4 Exercises

Exercise 1. Let u be a word. Show that u^* is star-free if and only if u is not of the form v^n, where v is a word and $n > 0$.

Exercise 2. Let $A = \{a, b\}$ and let L be the language $(ab + aba)^+$. Its minimal automaton is given below:



Compute the syntactic monoid M of L and show that M is aperiodic. Prove that L star-free and give a star-free expression for L.

Exercise 3. Let $A = \{a, b\}$ and let D_n be the sequence of languages defined by induction as follows: $D_0 = \emptyset^*$ and $D_{n+1} = (aD_nb)^*$. Prove that all the languages D_n are star-free. Give a star-free expression for D_2.

Exercise 4. The 2-shuffle of two languages K and L is the language

$$K \sqcup\!\sqcup_2 L = \{u_1v_1u_2v_2 \mid u_1, v_1, u_2, v_2 \in A^*, u_1u_2 \in K, v_1v_2 \in L\}.$$

(1) Give a rational expression for the language $(ab)^* \sqcup\!\sqcup_2 a^*$.
(2) Show that the 2-shuffle of two rational languages is rational.
(3) Show that the 2-shuffle of two star-free languages is star-free.

5 Notes

Schützenberger's theorem on star-free languages was published in [141]. It is usually considered as the root of algebraic automata theory and it had a considerable influence on subsequent research.

Proving that star-free languages are recognised by aperiodic monoids is the easiest part of the result. In his original proof, Schützenberger gave a direct construction to compute a monoid recognising the product of two languages, given two monoids recognising the given languages. This construction, nowadays called the *Schützenberger product of monoids*, is intimately related to the concatenation product (see [108] for an overview). With this tool in hand, it is easy to see that the product of two languages recognised by aperiodic monoids is also recognised by an aperiodic monoid. In this chapter, we followed another approach based on relational morphisms (see Chapter XVI, first introduced by Straubing [155, 154].

The most difficult part of Schützenberger's theorem consists in establishing that every language recognised by an aperiodic monoid is star-free. There are essentially two known ways to prove this result. We followed Schützenberger's original proof, which relies on the properties of Green's relations. Diekert and Kufleitner [40] recently proposed a new proof based on local divisors.

The alternative road, presented in Section XIX.5.3, relies on the wreath product principle. This approach was originally proposed by Cohen and Brzozowski [33] and by Meyer [88].

Chapter VII

Piecewise testable languages

Simon's theorem shows that the languages recognised by $\mathcal{J}$-trivial monoids are exactly the shuffle ideals. This result has far reaching consequences, both in semigroup theory and in automata theory.

As a preliminary step, we shall explore the properties of the subword ordering and give an algebraic characterisation of the shuffle ideals.

1 Subword ordering

Let A be a finite alphabet. Recall that a word $u = a_1 \dots a_k \in A^*$ (where $a_1, \dots, a_k$ are letters) is a *subword* of a word $v \in A^*$ it there exist words $v_0, v_1, \dots, v_k \in A^*$ such that $v = v_0 a_1 v_1 \cdots a_k v_k$. One also says that v is a *superword* of u. For instance, *ardab* is a subword of *abracadabra*.

The subword ordering is a partial ordering on A^*, which is compatible with the concatenation product. Here is another important property of the subword ordering, due to Higman:

Theorem 1.1. *A set of words of A^* that are pairwise incomparable for the subword ordering is necessarily finite.*

Proof. A sequence of words $(u_n)_{n \geqslant 0}$ is said to be *subword-free* if, for all $i < j$, u_i is not a subword of u_j. We claim there exist no infinite subword-free sequence. Otherwise, one would be able to find an "earliest" subword-free sequence, in the following sense:

(1) u_0 is a shortest word beginning a subword-free sequence of words,

(2) u_1 is a shortest word such that u_0, u_1 is the beginning of a subword-free sequence of words,

(3) u_2 is a shortest word such that u_0, u_1, u_2 is the beginning of a subword-free sequence of words, and so on.

Since A is finite, there exist infinitely many u_i that begin with the same letter a, say $u_{i_0} = av_{i_0}$, $u_{i_1} = av_{i_1}$, ..., with $i_0 < i_1 < \dots$ Let us show that the sequence

$$u_0, u_1, \dots, u_{i_0 - 1}, v_{i_0}, v_{i_1}, \dots \tag{1.1}$$

is subword-free. First of all, the sequence $u_0, u_1, \ldots, u_{i_0-1}$ is subword-free. Next, the sequence $v_{i_0}, v_{i_1}, \ldots$ is subword-free: if $i_r < i_s$, then u_{i_r} is not a subword of u_{i_s} and hence v_{i_r} is not a subword of v_{i_s}. Finally, if $0 \leqslant k < i_0$ and $r \geqslant 0$, then u_k is not a subword of v_{i_r}, for otherwise it would be a subword of u_{i_r}. Now, since v_{i_0} is shorter u_{i_0}, the sequence (1.1) is "earlier" that our original sequence, a contradiction. This proves the claim and the theorem follows. □

For each $n \geqslant 0$, we define an equivalence relation $\sim_n$ on A^* by $u \sim_n v$ if and only if u and v have the same subwords of length $\leqslant n$. For instance, $abbac \sim_1 cab$, since these two words have the same letters a, b and c, and $ababab \sim_3 bababa$ since any word of length $\leqslant 3$ is a subword of both words.

Proposition 1.2. *The relation $\sim_n$ is a congruence of finite index on A^*.*

Proof. Suppose that $u \sim_n v$ and let x, y be two words of A^*. Let w be a subword of xuy of length $\leqslant n$. The word w can be factored as $w_0w_1w_2$ where w_0, w_1 and w_2 are subwords of x, u and y, respectively. Since w_1 is shorter that w, $|w_1| \leqslant n$ and thus w_1 is also a subword of v. It follows that $w_0w_1w_2$ is a subword of xvy. Dually, every subword of xvy of length $\leqslant n$ is a subword of xuy. Thus $xuy \sim_n xvy$, showing that $\sim_n$ is a congruence.

The $\sim_n$-class of u is entirely determined by the set of subwords of u of length $\leqslant n$. Since there are finitely many such words, the congruence $\sim_n$ has finite index. □

We shall now establish some useful properties of this congruence.

Proposition 1.3. *Let $u, v \in A^*$ and $a \in A$. If $uav \sim_{2n-1} uv$, then either $ua \sim_n u$ or $av \sim_n v$.*

Proof. Suppose that $ua \not\sim_n u$ and $av \not\sim_n v$. Then there exists a word x of length $\leqslant n$ which is a subword of ua but not of u. Likewise there exists a word y of length $\leqslant n$ which is a subword of av but not of v. Necessarily one has $x = x'a$ and $y = ay'$ and $x'ay'$ is a word of length $\leqslant 2n-1$ which is a subword of uav but not of uv. Therefore $uav \not\sim_{2n-1} uv$. □

If u is a word, we let $c(u)$ denote the *content* of u, that is, the set of letters of A occurring in u. For instance, $c(babaa) = \{a, b\}$.

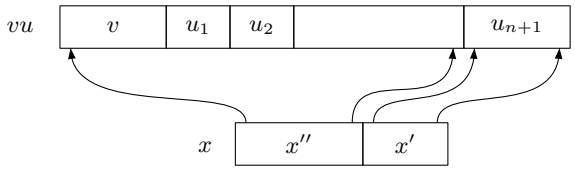
Proposition 1.4. *Let $u, v \in A^*$ and let $n > 0$. Then $u \sim_n vu$ if and only if there exist $u_1, \ldots, u_n \in A^*$ such that $u = u_1 \cdots u_n$ and $c(v) \subseteq c(u_1) \subseteq \ldots \subseteq c(u_n)$.*

Proof. First of all, the result is trivial if $u = 1$. We shall suppose from now on that u is nonempty.

Let us show that the condition is necessary by induction on n. If $n = 1$, $u \sim_1 vu$ implies that u and vu have the same content and hence $c(v) \subseteq c(u)$. Suppose that $u \sim_{n+1} vu$ and let u_{n+1} be the shortest suffix of u such that $c(u_{n+1}) = c(u)$. Since u is nonempty, so is u_{n+1}. Put $u_{n+1} = au'$ with $a \in A$. By definition of u_{n+1}, $c(u')$ is strictly contained in $c(u)$ and thus a is not a letter of u'. We claim that $w \sim_n vw$, where w is the prefix of u such that $u = wau'$. Let x be a subword of vw of length $\leqslant n$. Then xa is a subword of length $\leqslant n+1$ of vwa and therefore of vu. Since $u \sim_{n+1} vu$, xa is a subword of $u = wau'$ and, since a is not a letter of u', xa is a subword of wa. Therefore x is a subword

of w. Conversely, it is clear that every subword of w is a subword of vw, which proves the claim. By the induction hypothesis, there exist $u_1, \ldots, u_n \in A^*$ such that $w = u_1 \cdots u_n$ and $c(v) \subseteq c(u_1) \subseteq \ldots \subseteq c(u_n)$. Now $u = wu_{n+1}$ and $c(u_n) \subseteq c(u) = c(u_{n+1})$, which concludes the induction step.

We now show that the condition is sufficient, again by induction on n. For $n = 1$, $u_1 = u$ and $c(v) \subseteq c(u)$ implies $c(u) = c(vu)$, that is, $u \sim_1 vu$. Suppose that $u = u_1 \cdots u_{n+1}$ with $c(v) \subseteq c(u_1) \subseteq \ldots \subseteq c(u_{n+1})$. Then $c(vu) = c(u) = c(u_{n+1})$ and $u_1 \cdots u_n \sim_n vu_1 \cdots u_n$ by the induction hypothesis. Let x be a nonempty subword of length $\leqslant n + 1$ of vu. Let x' be the longest suffix of x such that x' is a subword of u_{n+1} and put $x = x''x'$.



Since $c(vu) = c(u_{n+1})$, the factor u_{n+1} contains each letter of vu, and hence of x, at least once. In particular, x' is nonempty. Furthermore, by the definition of x', x'' is a subword of length $\leqslant n$ of $vu_1 \cdots u_n$. Since $u_1 \cdots u_n \sim_n vu_1 \cdots u_n$, x'' is a subword of $u_1 \cdots u_n$ and therefore x is a subword of u. Consequently, every subword of u is a subword of vu and therefore $u \sim_{n+1} vu$, which completes the proof. □

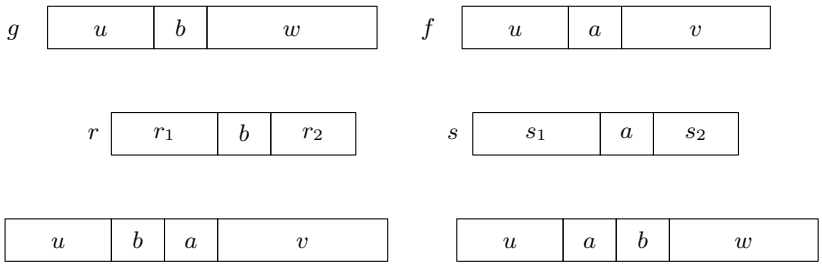
Corollary 1.5. *For every $u, v \in A^*$, one has $(uv)^n u \sim_n (uv)^n \sim_n v(uv)^n$.*

Proof. The formula $(uv)^n \sim_n v(uv)^n$ follows from Proposition 1.4. The other part of the formula is dual. □

We conclude this section with a remarkable combinatorial property of the congruence $\sim_n$.

Proposition 1.6. *If $f \sim_n g$, there exists h such that f and g are subwords of h and $f \sim_n h \sim_n g$.*

Proof. The proof is achieved by induction on $k = |f| + |g| - 2|f \wedge g|$ where $f \wedge g$ is the largest common prefix of f and g. If $k = 0$, then $f = g$ and it suffices to take $h = f = g$. The result is also trivial if f is a subword of g (or g is a subword of f). These cases are excluded from now on. Thus one has $f = uav$, $g = ubw$ with $a, b \in A$ and $a \neq b$. We claim that either $ubw \sim_n ubav$ or $uav \sim_n uabw$. Suppose that none of these assertions is true. Since $ubw = g \sim_n f$ and f is a subword of $ubav$, there exists a word r of length $\leqslant n$ which is a subword of $ubav$ but not of ubw. Likewise, there exists a word s of length $\leqslant n$ which is a subword of $uabw$ but not of uav.

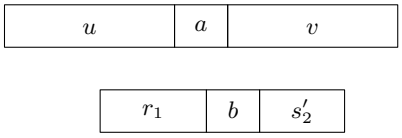


Necessarily $r = r_1 b r_2$ where r_1 is a subword of u and r_2 is a subword of av, and $s = s_1 a s_2$ where s_1 is a subword of u and s_2 is a subword of bw. It follows that $r_1 b$ is not a subword of u (for otherwise $r = r_1 b r_2$ would be a subword of $uav = f$ and therefore of g). Likewise $s_1 a$ is not a subword of u.

Since r_2 is a subword of av, one has $r_2 = r_2'' r_2'$ where $r_2'' = 1$ or a and r_2' is a subword of v. Likewise, since s_2 is a subword of bw; one has $s_2 = s_2'' s_2'$ where $s_2'' = 1$ or b and s_2' is a subword of w. Finally

$$|r_1 b s_2'| + |s_1 a r_2'| \leqslant |r_1 a s_2| + |s_1 b r_2| \leqslant |r| + |s| \leqslant 2n$$

and therefore one of the words $r_1 b s_2'$ or $s_1 a r_2'$ is of length $\leqslant n$. Suppose for example that this is $r_1 b s_2'$. Then $r_1 b s_2'$ is a subword of $ubw = g$ and therefore also of $f = uav$. However, $r_1 b$ is not a subword of u. Thus $b s_2'$ is a subword of v, and a fortiori s_2 is a subword of v.



Thus $s = s_1 a s_2$ is a subword of $uav = f$, a contradiction. This proves the claim. Suppose, for example, that $f = uav \sim_n uabw$. Then

$$\begin{aligned} |uav| + |uabw| - 2|uav \wedge uabw| &\leqslant |f| + |g| + 1 - 2|ua| \\ &\leqslant |f| + |g| + 1 - (2|f \wedge g| + 2) \\ &< k \end{aligned}$$

By the induction hypothesis, there exists h such that $f = uav$ is a subword of h, $uabw$ is a subword of h and $f \sim_n h \sim_n uabw$. The proposition follows from this, since g is a subword of $uabw$. □

Example 1.1. Let $f = a^3 b^3 a^3 b^3$ and $g = a^2 b^4 a^4 b^2$. We have $f \sim_4 g$ since all words of length 4 except $baba$ are subwords of f and g. Applying the algorithm described in the proof of Proposition 1.6, we obtain successively

$$f = (aa)a(b^3 a^3 b^3) \sim_4 (aa)b(b^3 a^4 b^2) = g$$

whence

$$(aa)a(b^3 a^3 b^3) \sim_4 (aa)ab(b^3 a^4 b^2) \quad \text{or} \quad (aa)b(b^3 a^4 b^2) \sim_4 (aa)ba(b^3 a^3 b^3)$$

The second possibility can be ruled out, for $baba$ is a subword of $a^2bab^3a^3b^3$. Therefore

$$(a^3b^3)a(a^2b^3) \sim_4 (a^3b^3)b(a^4b^2)$$

and consequently

$$(a^3b^3)a(a^2b^3) \sim_4 (a^3b^3)ab(a^4b^2) \quad \text{or} \quad (a^3b^3)b(a^4b^2) \sim_4 (a^3b^3)ba(a^2b^3)$$

The first possibility can be ruled out, for $baba$ is a subword of $a^3b^3aba^4b^2$. Then

$$(a^3b^4a^3)a(b^2) \sim_4 (a^3b^4a^3)b(b^2)$$

and consequently

$$(a^3b^4a^3)a(b^2) \sim_4 (a^3b^4a^3)ab(b^2) \quad \text{or} \quad (a^3b^4a^3)b(b^2) \sim_4 (a^3b^4a^3)ba(b^2)$$

The second possibility can be ruled out, for $baba$ is a subword of $a^3b^4a^3bab^2$. Therefore

$$a^3b^4a^4b^2 \sim_4 a^3b^4a^4b^3$$

It follows from this that f and g are subwords of $h = a^3b^4a^4b^3$ and that $f \sim_4 h \sim_4 g$.

2 Simple languages and shuffle ideals

The *shuffle* of two languages L_1 and L_2 of A^* is the language $L_1 \sqcup\!\sqcup L_2$ of A^* defined by:

$$L_1 \sqcup\!\sqcup L_2 = \{w \in A^* \mid w = u_1v_1 \cdots u_nv_n \text{ for some } n \geqslant 0 \text{ such that } u_1 \cdots u_n \in L_1, v_1 \cdots v_n \in L_2\}$$

In particular, if L is a language of A^*, a language of the form $L \sqcup\!\sqcup A^*$ is called a *shuffle ideal*. Thus a language L of A^* is a shuffle ideal if every superword of a word of L is also in L.

A *simple language* is a shuffle ideal of the form

$$A^* \sqcup\!\sqcup a_1 \dots a_k = A^*a_1A^*a_2A^* \cdots A^*a_kA^*$$

where $a_1, \dots, a_k \in A$. Thus $A^*a_1A^*a_2A^* \cdots A^*a_kA^*$ is the set of superwords of the word $a_1 \cdots a_k$. We can now state our first characterisation of shuffle ideals:

Theorem 2.7. *A language is a shuffle ideal if and only if it is a finite union of simple languages.*

Proof. Clearly, every finite union of simple languages is a shuffle ideal. Conversely, let L be a shuffle ideal and let F be the set of all minimal words of L for the subword ordering. Thus L is the set of all superwords of F, that is $L = F \sqcup\!\sqcup A^*$. Furthermore, since the elements of F are pairwise incomparable for the subword ordering, Higman's theorem (Theorem 1.1) shows that F is finite. Therefore L is the finite union of the simple languages $A^* \sqcup\!\sqcup u$, where the union runs over all words $u \in F$. □

Corollary 2.8. *Every shuffle ideal is a recognisable language.*

One can give a constructive proof which does not rely on Higman's theorem.

Proposition 2.9. *Let L be recognisable language such that $L \sqcup A^* = L$. Then one can effectively find a finite language F such that $L = F \sqcup A^*$.*

Proof. Let n be the number of states of the minimal automaton $\mathcal{A}$ of L. Set

$$F = \{u \in L \mid |u| \leqslant n\} \text{ and } K = F \sqcup A^*$$

We claim that $L = K$. Since $F \subseteq L$, one has $F \sqcup A^* \subseteq L \sqcup A^* = L$ and hence $K \subseteq L$. If the inclusion is strict, consider a word u of minimal length in $L - K$. Necessarily, $|u| > n$, for otherwise $u \in F$. Let $u = a_1 \cdots a_r$ and let $q_0 \xrightarrow{a_1} q_1 \cdots q_{r-1} \xrightarrow{a_r} q_r$ be a successful path with label u in $\mathcal{A}$. As $r > n$, there exist two indices $i < j$ such that $q_i = q_j$. Thus the word $v = a_1 \cdots a_i a_{j+1} \cdots a_r$ is also accepted by $\mathcal{A}$ and therefore belongs to L. Furthermore, since v is shorter than u, v belongs to K and u belongs to $K \sqcup A^*$. Now, since

$$K \sqcup A^* = (F \sqcup A^*) \sqcup A^* = F \sqcup (A^* \sqcup A^*) = F \sqcup A^* = K$$

one has $u \in K$, a contradiction. This proves the claim and the proposition. □

We now come to the algebraic characterisation of shuffle ideals. Let us say that an ordered monoid M satisfies the identity $1 \leqslant x$ if, for all $x \in M$, $1 \leqslant x$.

Theorem 2.10. *A language is a shuffle ideal if and only if its syntactic ordered monoid satisfies the identity $1 \leqslant x$.*

Proof. Let L be a language and let $\eta : A^* \to (M, \leqslant_L)$ be its syntactic ordered morphism. Suppose that L is a shuffle ideal. If $uv \in L$, then $uxv \in L$ for each $x \in A^*$. Therefore $1 \leqslant_L x$ and thus M satisfies the identity $1 \leqslant x$.

Conversely, if $(M, \leqslant_L)$ satisfies the identity $1 \leqslant x$, then, for every $x \in A^*$, $1 \leqslant_L x$, that is, the condition $uv \in L$ implies $uxv \in L$. Therefore L is a shuffle ideal. □

3 Piecewise testable languages and Simon's theorem

A language is called *piecewise testable* if and only if it is a union of $\sim_n$-classes for some positive integer n.

The terminology chosen can be explained as follows: a language L is piecewise testable if there exists an integer $n > 0$ such that one can test whether or not a word belongs to L by simple inspection of its subwords of length $\leqslant n$. Here is a first description of these languages.

Proposition 3.11. *A language of A^* is piecewise testable if and only if it belongs to the Boolean algebra generated by the simple languages on A^*.*

Proof. Let $L = A^* a_1 A^* \cdots a_n A^*$ be a simple language of A^*. If $u \in L$, then $a_1 \cdots a_n$ is a subword of u. Therefore, if $u \sim_n v$, $a_1 \cdots a_n$ is also a subword of v and $v \in L$. This shows that L is saturated by $\sim_n$ and therefore is a finite union of $\sim_n$-classes.

Let u be a word of A^*. A moment's reflection should suffice to verify the following formula:

$$\{v \in A^* \mid v \sim_n u\} = \Big(\bigcap_{a_1 \cdots a_k \in E} A^* a_1 A^* \cdots a_k A^* \Big) - \Big(\bigcup_{a_1 \cdots a_k \in F} A^* a_1 A^* \cdots a_k A^* \Big)$$

where E is the set of subwords of u of length $\leqslant n$ and F is the set of words of length $\leqslant n$ which are not subwords of u. It follows from this formula that if L is a union of $\sim_n$-classes for some positive integer n, then L belongs to the Boolean algebra generated by the simple languages on A^*. □

The syntactic characterisation of piecewise testable languages is the main result of this chapter. It relies on two results of semigroup theory that have independent interest.

Proposition 3.12. *Every finite ordered monoid satisfying the identity $1 \leqslant x$ is $\mathcal{J}$-trivial.*

Proof. Let x and y be two elements of M such that $x \mathcal{J} y$. Then $x = rys$ and $y = uxv$ for some $r, s, u, v \in M$. Since $1 \leqslant u$ and $1 \leqslant v$, it follows that $x \leqslant uxv = y$ and similarly, $y \leqslant x$. Thus $x = y$. □

Theorem 3.13 (Simon)**.** *Let M be a finite $\mathcal{J}$-trivial monoid and let n be the maximal length of strict $<_{\mathcal{J}}$-chains in M. If $\varphi : A^* \to M$ is a surjective morphism, then M is a quotient of the monoid $A^*/\sim_{2n-1}$.*

Proof. By Proposition II.3.22, it suffices to show that if $f \sim_{2n-1} g$, then $\varphi(f) = \varphi(g)$. By Proposition 1.6, we may assume that f is a subword of g. We note furthermore that if f is a subword of h and h is a subword of g, then we also have $f \sim_{2n-1} h$. This enables us to assume that $f = uv$ and $g = uav$ for some $a \in A$. In this case, Proposition 1.3 shows that either $ua \sim_n u$ or $av \sim_n v$. Assuming the latter, there exists by Proposition 1.4 a factorisation $v = v_1 v_2 \cdots v_n$ such that $\{a\} \subseteq c(v_1) \subseteq \ldots \subseteq c(v_n)$. Consider the $\leqslant_{\mathcal{J}}$-chain of length $n+1$

$$\varphi(v_1 \cdots v_n) \leqslant_{\mathcal{J}} \varphi(v_2 \cdots v_n) \leqslant_{\mathcal{J}} \cdots \leqslant_{\mathcal{J}} \varphi(v_n) \leqslant_{\mathcal{J}} 1$$

By the choice of n, this chain is not strict and there exists an index i such that $\varphi(v_i \cdots v_n) \mathcal{J} \varphi(v_{i+1} \cdots v_n)$. Since M is $\mathcal{J}$-trivial, one has $\varphi(v_i \cdots v_n) = \varphi(v_{i+1} \cdots v_n) = s$. Let $b \in c(v_i)$. Then $v_i = v_i' b v_i''$ for some $v_i', v_i'' \in A^*$ and thus $s = \varphi(v_i \cdots v_n) \leqslant_{\mathcal{J}} \varphi(b v_i'' v_{i+1} \cdots v_n) \leqslant_{\mathcal{J}} \varphi(v_i'' v_{i+1} \cdots v_n) \leqslant_{\mathcal{J}} \varphi(v_{i+1} \cdots v_n) = s$. Since M is $\mathcal{J}$-trivial, we get

$$s = \varphi(v_i \cdots v_n) = \varphi(b v_i'' v_{i+1} \cdots v_n) = \varphi(v_i'' v_{i+1} \cdots v_n) = \varphi(v_{i+1} \cdots v_n).$$

Consequently, one has $\varphi(b)s = s$ for each $b \in c(v_i)$ and hence also for $b \in c(v_{i-1})$, $\ldots$, $b \in c(v_0)$ and $b = a$. Therefore $\varphi(v) = \varphi(v_1 \cdots v_n) = s = \varphi(a)s = \varphi(av)$. It follows that $\varphi(g) = \varphi(uav) = \varphi(uv) = \varphi(f)$, which concludes the proof. □

We now return to the announced characterisation of piecewise testable languages.

Theorem 3.14 (Simon)**.** *A language is piecewise testable if and only if its syntactic monoid is finite and $\mathcal{J}$-trivial.*

Proof. Let L be a simple language. Then by Theorem 2.10, the syntactic ordered monoid of L satisfies the identity $1 \leqslant x$. By Proposition 3.12, this monoid is $\mathcal{J}$-trivial. Now if L is piecewise testable, it is by Proposition 3.11 a Boolean combination of simple languages, its syntactic monoid divides a product of finite $\mathcal{J}$-trivial monoids and hence is itself finite and $\mathcal{J}$-trivial.

Conversely, if the syntactic monoid of L is finite and $\mathcal{J}$-trivial, then by Theorem 3.13, L is a union of $\sim_{2n-1}$-classes, where n is the maximal length of strict $<_{\mathcal{J}}$-chains in M. Thus L is piecewise testable. □

4 Some consequences of Simon's theorem

Simon's theorem has unexpected consequences in semigroup theory. We start by defining, for each integer $n > 0$, three monoids $\mathcal{C}_n$, $\mathcal{R}_n$ and $\mathcal{U}_n$ which will serve us as examples of $\mathcal{J}$-trivial monoids.

The monoid $\mathcal{C}_n$ is the submonoid of $\mathcal{T}_n$ consisting of all order preserving and extensive functions from $\{1, \ldots, n\}$ to itself. Recall that a transformation a on $\{1, \ldots, n\}$ is *order preserving* if $p \leqslant q$ implies $p \cdot a \leqslant q \cdot a$ and *extensive* if for all p, $p \leqslant p \cdot a$.

The monoid $\mathcal{R}_n$ is the monoid of all reflexive relations on $\{1, \ldots, n\}$. It is convenient to consider $\mathcal{R}_n$ as the monoid of Boolean matrices of size $n \times n$ having only entries 1 on the diagonal. For example

$$\mathcal{R}_2 = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \right\}$$

Finally, $\mathcal{U}_n$ is the submonoid of $\mathcal{R}_n$ consisting of the upper triangular matrices of $\mathcal{C}_n$. The matrices of $\mathcal{U}_n$ are called *unitriangular*. For example,

$$\mathcal{U}_3 = \left\{ \begin{pmatrix} 1 & \varepsilon_1 & \varepsilon_2 \\ 0 & 1 & \varepsilon_3 \\ 0 & 0 & 1 \end{pmatrix} \mid \varepsilon_1, \varepsilon_2, \varepsilon_3 \in \{0, 1\} \right\}$$

Proposition 4.15. *For each $n > 0$, the monoids $\mathcal{C}_n$, $\mathcal{R}_n$ and $\mathcal{U}_n$ are $\mathcal{J}$-trivial.*

Proof. Let us show that $\mathcal{C}_n$ is $\mathcal{J}$-trivial. If $f, g \in \mathcal{C}_n$ and $f \mathrel{\mathcal{J}} g$, then $g = afb$ and $f = cgd$ for some $a, b, c, d \in \mathcal{C}_n$. Let $p \in \{1, \ldots, n\}$. Since a is extensive, one has $p \leqslant p \cdot a$ and since f is order-preserving, one has $p \cdot f \leqslant p \cdot af$. It follows, since b is extensive, that $p \cdot af \leqslant p \cdot afb$ and finally $p \cdot f \leqslant p \cdot afb = p \cdot g$. Similar reasoning would show that $p \cdot g \leqslant p \cdot f$. It follows that $f = g$ and thus $\mathcal{C}_n$ is $\mathcal{J}$-trivial.

Since $\mathcal{U}_n$ is a submonoid of $\mathcal{R}_n$, it is sufficient to establish that $\mathcal{R}_n$ is $\mathcal{J}$-trivial. But $\mathcal{R}_n$ is naturally ordered by the order defined by $m \leqslant n$ if and only if, for all i, j, $m_{i,j} \leqslant n_{i,j}$ and this order is stable under product. Since all entries on the diagonal are equal to 1, the identity $1 \leqslant x$ holds in $\mathcal{R}_n$ and thus $\mathcal{R}_n$ is $\mathcal{J}$-trivial by Proposition 3.12. □

The next proposition gives another property of the monoids $\mathcal{C}_n$, $\mathcal{R}_n$ and $\mathcal{U}_n$.

Proposition 4.16. *For each $n, m > 0$, the monoid $\mathcal{C}_n \times \mathcal{C}_m$, $[\mathcal{R}_n \times \mathcal{R}_m, \mathcal{U}_n \times \mathcal{U}_m]$ is isomorphic to a submonoid of $\mathcal{C}_{n+m}$ $[\mathcal{R}_{n+m}, \mathcal{U}_{n+m}]$.*

Proof. Let $\varphi : \mathcal{C}_n \times \mathcal{C}_m \to \mathcal{C}_{n+m}$ be the function defined by $\varphi(f, g) = h$ where

$$p \cdot h = \begin{cases} p \cdot f & \text{if } 1 \leqslant p \leqslant n, \\ (p - n) \cdot g + n & \text{if } n + 1 \leqslant p \leqslant n + m \end{cases}$$

Then φ is clearly an injective morphism and therefore $\mathcal{C}_n \times \mathcal{C}_m$ is isomorphic to a submonoid of $\mathcal{C}_{n+m}$.

Let now $\psi : \mathcal{R}_n \times \mathcal{R}_m \to \mathcal{R}_{n+m}$ be the function defined by $\psi(R, S) = T$ where T is the relation defined by $(i, j) \in T$ if and only if $(i, j) \in R$ or $(i-n, j-n) \in S$. Then ψ is an injective morphism and therefore $\mathcal{R}_n \times \mathcal{R}_m$ is isomorphic to a submonoid of $\mathcal{R}_{n+m}$. The proof is similar for $\mathcal{U}_n$. □

The next result shows that the monoids $\mathcal{C}_n$, $\mathcal{R}_n$ and $\mathcal{U}_n$ generate the variety of $\mathcal{J}$-trivial monoids.

Theorem 4.17. *Let M be a finite monoid. The following conditions are equivalent:*

(1) *M is $\mathcal{J}$-trivial,*
(2) *there exists an integer $n > 0$ such that M divides $\mathcal{C}_n$,*
(3) *there exists an integer $n > 0$ such that M divides $\mathcal{R}_n$,*
(4) *there exists an integer $n > 0$ such that M divides $\mathcal{U}_n$.*

Proof. By Proposition 4.15, the monoids $\mathcal{C}_n$, $\mathcal{R}_n$ and $\mathcal{U}_n$ are $\mathcal{J}$-trivial. Therefore each of the conditions (2), (3) or (4) implies (1). Moreover (4) implies (3) since $\mathcal{U}_n$ is a submonoid of $\mathcal{R}_n$. It remains to prove that (1) implies (2) and (4).

Let M be a $\mathcal{J}$-trivial monoid. By Proposition XIII.4.8, there exist a finite alphabet A and languages $L_1, \ldots, L_k \in \mathcal{J}(A^*)$ such that M is isomorphic to a submonoid of $M(L_1) \times \cdots \times M(L_k)$. Now by Theorem 3.14 and by Proposition 3.11, each L_i is a Boolean combination of simple languages. It follows now from Proposition IV.4.27 that M divides a product of syntactic monoids of simple languages. Therefore, by Proposition 4.16, it suffices to establish (2) and (4) when M is the syntactic monoid of a simple language $L = A^* a_1 A^* \cdots a_n A^*$. The minimal automaton $\mathcal{A}$ of L is pictured in Figure 4.1:

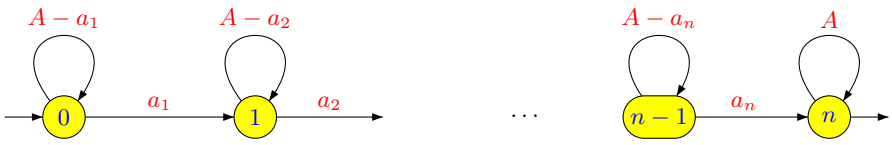


Figure 4.1. An automaton recognising L.

Since the transitions of $\mathcal{A}$ are increasing and extensive functions, the transition monoid of $\mathcal{A}$, which is also the syntactic monoid of L, is a submonoid of $\mathcal{C}_{n+1}$, which proves that (1) implies (2).

Furthermore, L is also recognised by the nondeterministic automaton $\mathcal{B}$ represented in Figure 4.2:

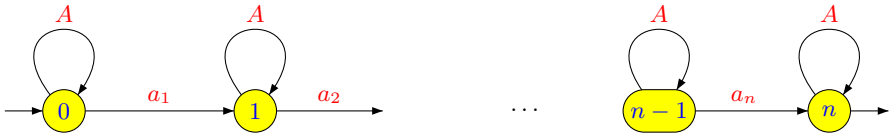


Figure 4.2. A nondeterministic automaton recognising L.

The transition monoid of $\mathcal{B}$ is a monoid of unitriangular Boolean matrices, that is, a submonoid of $\mathcal{U}_{n+1}$. It follows that M divides $\mathcal{U}_{n+1}$. □

Theorem 3.13 has another very important consequence.

Theorem 4.18. *Every finite $\mathcal{J}$-trivial monoid is a quotient of a finite ordered monoid satisfying the identity $1 \leqslant x$.*

Proof. Let M be a $\mathcal{J}$-trivial monoid. By Theorem 3.13, M is a quotient of the monoid of the form $A^*/\sim_k$. Now, the subword ordering induces a stable partial order on $A^*/\sim_k$. Furthermore, since the empty word is a subword of every word, the identity $1 \leqslant x$ holds in this ordered monoid. □

5 Exercises

Exercise 1. Let $L = A_0^* a_1 A_1^* \cdots a_n A_n^*$, where $a_1, \ldots, a_n \in A$, $A_0, \ldots, A_n \subseteq A$ and $a_1 \notin A_0 \cup A_1$, $a_2 \notin A_1 \cup A_2$, $\ldots$, $a_n \notin A_{n-1} \cup A_n$. Show that L is piecewise testable. Use this result to show that the language $abab^*ca^*b$ is piecewise testable.

Exercise 2. Let L be a language. Prove that the following conditions are equivalent:

(1) L is the complement of a shuffle ideal,
(2) L is a finite union of languages of the form $A_0^*(1 + a_1)A_1^* \cdots (1 + a_n)A_n^*$, where $a_1, \ldots, a_n \in A$ and $A_0, \ldots, A_n \subseteq A$,
(3) L satisfies the identity $x \leqslant 1$.

Exercise 3. Let A be an alphabet. A language L of A^* is said to be *stutter-invariant* if, for each letter $a \in A$, $a \sim_L a^2$. The aim of this problem is to study the class $\mathcal{C}$ of all languages that are both stutter-invariant and piecewise testable.

A language of the form

$$A^* a_1 A^* a_2 \cdots A^* a_k A^*$$

where the a_i are letters such that $a_i \neq a_{i+1}$, for $1 \leqslant i \leqslant k-1$, is said to be *elementary stutter-invariant*.

(1) Prove that any langage elementary stutter-invariant is stutter-invariant.
(2) Prove that any Boolean combination of elementary stutter-invariant languages belongs to $\mathcal{C}$.
(3) (More difficult) Prove that, conversely, each language of $\mathcal{C}$ is a Boolean combination of elementary stutter-invariant languages.

(4) (This question requires the notion of profinite equations introduced in Chapter XII). Give a set of profinite equations defining the class $\mathcal{C}$.

6 Notes

We followed Simon's original proof [145], which is based on a careful analysis of the subword ordering and has a strong combinatorial flavour. Several other proofs of Simon's theorem are known. A nice proof via combinatorics on words was proposed by Klíma [70]. Straubing and Thérien [163] gave a purely algebraic proof by proving directly Theorem 4.18. Another direct proof of 4.18 is due to Henckell and Pin [57]. Almeida [3] proposed an approach using profinite words.

Chapter VIII

Locally testable languages

1 Scanners

In this section, we are interested in a special class of automata, called scanners. Scanners can be considered as a model for computation that require only "local" information. Informally, a scanner is an automaton equipped with a finite memory and a "sliding" window of a fixed length. In a typical computation, the sliding window is moved from left to right on the input, so that the scanner can remember the factors of length smaller than or equal to the size of the window. In view of these factors, the scanner decides whether or not the input is accepted or rejected.

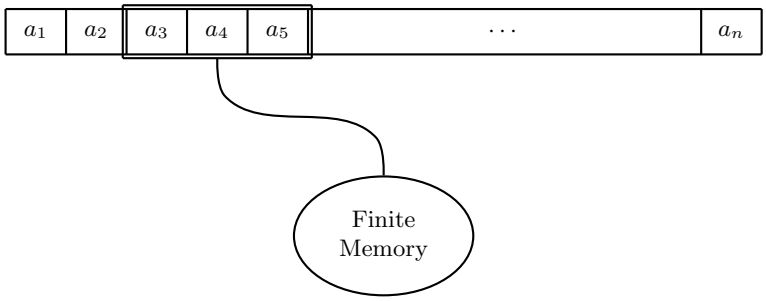


Figure 1.1. A scanner.

The sliding window, say of size k, is allowed to move beyond the first and the last letter of the word, so that the prefixes and suffixes of length $< k$ can be read. For instance, if $k = 3$, and $u = abbaaabab$, different positions of the window are represented in the following diagram:

$$\boxed{\ \ a}bbaaabab \quad \boxed{\ \ ab}baaabab \quad \boxed{abb}aaabab \quad a\boxed{bba}aabab \quad \cdots \quad abbaaaba\boxed{b\ \ }$$

Scanners can be used to recognise the local languages introduced in Section III.5.1. Roughly speaking, a local language is described by the factors of length 2 of its words. For instance, if $A = \{a, b, c, d\}$, the language $c(ab)^+d$ is the set of all words whose set of factors of length 2 is exactly $\{ca, ab, ba, bd\}$. The

locally testable languages generalize local languages: the membership of a given word in such a language is determined by the set of factors of a fixed length k (the order in which these factors occur and their frequency is not relevant) of the word, and by the prefixes and suffixes of length $< k$ of the word. These conditions can be tested by a scanner.

2 A congruence on words

For each positive integer k, let $\sim_k$ be the relation on A^* defined by $u \sim_k v$ if and only if

(a) u and v have the same prefixes of length $< k$,

(b) u and v have the same suffixes of length $< k$,

(c) u and v have the same factors of length k (without counting multiplicities).

For instance, $ababab cbcb \sim_3 ababcbcbcb$.

Proposition 2.1. *For each positive integer k, the relation $\sim_k$ is a congruence of finite index.*

Proof. It is clear that $\sim_k$ is an equivalence relation. It is also a congruence since if $u \sim_k v$ and a is a letter, then $au \sim_k av$ and $ua \sim_k va$. Finally, $\sim_k$ has finite index since the equivalence classes depend only of the following parameters: the prefixes of length $< k$, the suffixes of length $< k$ and the factors of length k. In each case there are only finitely many possible choices. □

A language is *k-testable* if it is union of $\sim_k$-classes. It is *locally testable* if it is k-testable for some k.

Proposition 2.2. *A language is locally testable if and only if it is a Boolean combination of languages of the form pA^*, A^*uA^* or A^*s with $p, u, s \in A^+$..*

Proof. Let $k = |p|$. If $u \in pA^*$ and $u \sim_k v$ then p is a prefix of v and thus $v \in pA^*$. It follows that pA^* is locally testable. A similar argument would show that A^*uA^* and A^*s are locally testable.

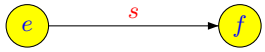
Let $x \in A^+$. If $|x| < k$, the $\sim_k$-class of x is $\{x\}$, which can be written as $xA^* - \bigcup_{a\in A} A^*xaA^*$. If $|x| \geqslant k$, let p [s] be its prefix [suffix] of length $k-1$ and let F be the set of its factors of length k. Then the $\sim_k$-class of x is the set

$$pA^* \cap A^*s \cap \Big(\bigcap_{u\in F} A^*uA^* \setminus \bigcup_{u\in A^k - F} A^*uA^*\Big)$$

In all cases it is a Boolean combination of languages of the form pA^*, A^*uA^* or A^*s.

3 Path conditions

Let L be a regular language of A^+, let S be its syntactic semigroup (not monoid!) and let $\pi : A^+ \to S$ its syntactic morphism. Let us denote by $G(S)$ the directed graph whose vertices are the idempotents of S and the edges are of the form



where $e, f \in E(S)$ and s is an element of S such that $es = s = sf$.

Example 3.1. Consider the alphabet $A = \{a, b\}$ and the language $L = A^*abbA^*$, whose minimal automaton is given below:

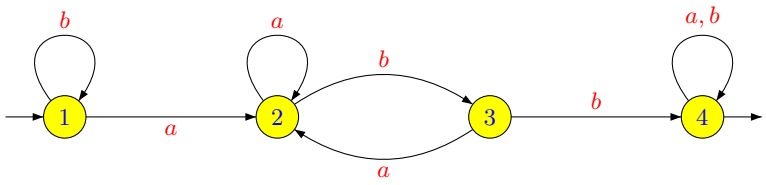


Figure 3.1. The minimal automaton of A^*abbA^*.

The elements of the syntactic monoid of L are

	1	2	3	4
$*\ 1$	1	2	3	4
$*\ a$	2	2	2	4
b	1	3	4	4
$*\ ab$	3	3	3	4
$*\ ba$	2	2	4	4
$*\ b^2$	1	4	4	4
$*\ ab^2$	4	4	4	4
bab	3	3	4	4
b^2a	2	4	4	4
b^2ab	3	4	4	4

Note that ab^2 is a zero of M. Thus we set $ab^2 = 0$. The other relations defining M are:

$$a^2 = a \qquad aba = a \qquad b^3 = b^2$$

Idempotents:

$$E(S) = \{1, a, ab, ba, b^2, ab^2\}$$

$\mathcal{D}$-classes:

$*\ 1$

b

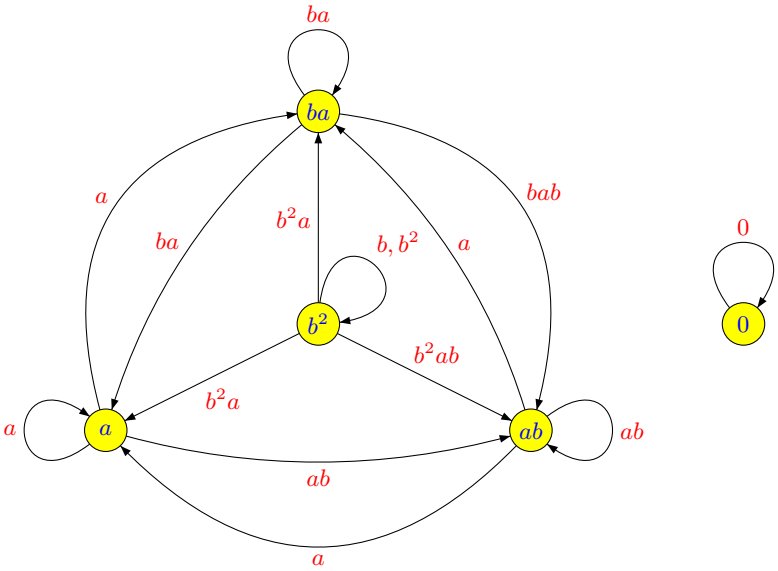
$*\ b^2$

$*\ a$	$*\ ab$
$*\ ba$	bab

b^2a	b^2ab

$^*ab^2$

Let $S = M - \{1\}$ be the syntactic semigroup of L. The graph $G(S)$ is partially represented in the picture below. The edges of the form $(e, 0, f)$ should be added.



By definition, the *label* of the path

of $G(S)$ is the product $s_1 s_2 \cdots s_n$.

A semigroup S satisfies the *path condition* if two paths of $G(S)$ with the same origin and the same end, and containing the same edges (without counting multiplicities), have the same labels.

4 An algebraic characterization

Locally testable languages are characterized by a deep algebraic property of their syntactic semigroup, discovered independently by Brzozowski-Simon [23] and McNaughton [85].

Let S be a finite semigroup. A local subsemigroup*local semigroup* of S is a subsemigroup of S of the form eSe, where $e \in E(S)$ (see Exercise II.10). A semigroup is said to be *locally trivial*, (respectively *locally commutative*, *locally idempotent*, *locally a group* locally a group, etc.) if all the local subsemigroups of S are trivial (respectively commutative, idempotent, groups, etc.). For instance, a semigroup S is locally idempotent and commutative if, for each $e \in E(S)$ and each $s, t \in S$, $(ese)^2 = (ese)$ and $(ese)(ete) = (ete)(ese)$.

Theorem 4.3. [23, 85, 158] *A language is locally testable if and only if its syntactic semigroup is finite and locally idempotent and commutative.*

TO DO.

5 Notes

Brzozowski and Simon[23] and independently McNaughton [85].

Scanners were first introduced in [12]. There are several possible variations on this definition. First, one can drop the conditions about the prefixes and suffixes. Membership in this type of language, that we call strongly locally testable (SLT), is determined only by factors of a fixed length k. Thus, a language is SLT if and only if it is a finite Boolean combination of languages of the form A^*wA^*, where w is a word. One can show that this family is also decidable and characterized by another nice algebraic property. But this time, the syntactic semigroup does not suffice, and a property of the image of the language in its syntactic semigroup is needed.

A second natural extension is to take in account the number of occurrences of the factors of the word. However, since we want to use finite automata to recognise our languages, we can only count factors up to a certain threshold. Threshold counting is the favorite way of counting of small children : they can distinguish 0, 1, 2, ... but after a certain number n (the threshold), all numbers are "big". From a more mathematical point of view, two positive integers s and t are congruent threshold n if $s = t$ or if $s \geqslant n$ and $t \geqslant n$. This defines the threshold locally testable languages (TLT). A combination of two deep results of Straubing [158] and Thérien and Weiss [167] yields a syntactic characterization of these languages. Similar results hold if one drops the conditions about the prefixes and suffixes.

Chapter IX

An excursion into logic

1 Introduction

The links between finite automata, languages and logic were first discovered by Büchi [24] and Elgot [43]. Their main result states that monadic second order has the same expressive power as finite automata.

The formalism of logic is introduced in Section 2. This section can be skipped by a reader with some background in logic. We detail the interpretation of formulas on words. This amounts mainly to considering a word as a structure by associating with each positive integer i the letter in position i. The relations that are used are the order relation between the indices and the relation $\mathbf{a}(i)$ expressing that the letter in position i is an a.

In Section 3, we prove the equivalence between finite automata and monadic second-order logic (Büchi's theorem). The conversion from automata to logic is obtained by writing a formula expressing that a word is the label of a successful path in an automaton. The opposite direction is more involved and requires to interpret formulas with free variables as languages over extended alphabets.

In Section 4, we present the corresponding theory for first-order logic. We prove McNaughton's theorem [86], stating that first-order logic of the linear order captures star-free languages.

2 The formalism of logic

In this section, we review some basic definitions of logic: first-order logic, second-order, monadic second-order and weak monadic second-order logic.

2.1 Syntax

Let us start by defining the syntax of *first-order logic*.

The basic ingredients are the *logical symbols* which encompass the logical connectives: $\wedge$ (and), $\vee$ (or), $\neg$ (not), $\rightarrow$ (implies), the equality symbol $=$, the *quantifiers* $\exists$ (there exists) and $\forall$ (for all), an infinite set of variables (most often denoted by x, y, z, or x_0, x_1, x_2,...) and parenthesis (to ensure legibility of the formulas).

In addition to these logical symbols, we make use of a set $\mathcal{L}$ of nonlogical symbols, called the *signature* of the first-order language. These auxiliary symbols can be of three types: relation symbols (for instance $<$), function symbols (for instance min), or constant symbols (for instance 0, 1). Expressions are built from the symbols of $\mathcal{L}$ by obeying the usual rules of the syntax, then first-order formulas are built by using the logical symbols, and are denoted by $\mathbf{FO}[\mathcal{L}]$. We now give a detailed description of the syntactic rules to obtain the logical formulas in three steps, passing successively, following the standard terminology, from terms to atomic formulas and subsequently to formulas.

We first define the set of $\mathcal{L}$-*terms*. It is the least set of expressions containing the variables, the constant symbols of $\mathcal{L}$ (if there are any) which is closed under the following formation rule: if $t_1, t_2, \ldots, t_n$ are terms and if f is a function symbol with n arguments, then the expression $f(t_1, t_2, \ldots, t_n)$ is a term. In particular, if $\mathcal{L}$ does not contain any function symbol, the only terms are the variables and the constant symbols.

Example 2.1. Let us take as set of nonlogical symbols

$$\mathcal{L} = \{ <, g, 0\}$$

where $<$ is a binary relation symbol, g is a two-variable function symbol and 0 is a constant. Then the following expressions are terms:

$$x_0 \qquad g(0,0) \qquad g(x_1, g(0, x_2)) \qquad g(g(x_0, x_1), g(x_1, x_2))$$

The *atomic formulas* are formulas either of the form

$$(t_1 = t_2)$$

where t_1 and t_2 are terms, or of the form

$$R(t_1, \ldots, t_n)$$

where $t_1, \ldots, t_n$ are terms and R is a n-ary relation symbol of $\mathcal{L}$.

Example 2.2. Continuing Example 2.1, the following expressions are atomic formulas:

$$(g(x_1, g(0, x_2)) = x_1) \qquad (g(0,0) < 0)$$
$$(g(g(x_0, x_1), g(x_1, x_2)) < g(x_3, x_1))$$

Notice that, in order to improve legibility, we did not apply literally the definition of atomic formulas: indeed, since $<$ is a symbol of binary relation, one should write $<(t_1, t_2)$ instead of $t_1 < t_2$.

Finally, the *first-order formulas* on $\mathcal{L}$ form the least set of expressions containing the atomic formulas and closed under the following formation rules:

(i) If $(\varphi_i)_{i \in I}$ is a finite family of first-order formulas, so are the expressions

$$\Big(\bigwedge_{i \in I} \varphi_i\Big) \quad \text{and} \quad \Big(\bigvee_{i \in I} \varphi_i\Big)$$

(ii) If φ and ψ are first-order formulas, so are the expressions

$$\neg\varphi \quad \text{and} \quad (\varphi \to \psi)$$

(iii) If φ is a first-order formula and if x is a variable, then the expressions

$$(\exists x\varphi) \quad \text{and} \quad (\forall x\varphi)$$

are first-order formulas.

To make notations easier, we set

$$\mathbf{true} = \bigwedge_{i\in\emptyset} \varphi_i \quad \text{and} \quad \mathbf{false} = \bigvee_{i\in\emptyset} \varphi_i$$

Example 2.3. The following expressions are first-order formulas of our example language:

$$(\exists x\ (\forall y\ ((y < g(z,0)) \wedge (x < 0)))) \qquad (\forall x\ (y = x))$$

Again, it is convenient to simplify notation by suppressing some of the parenthesis and we shall write the previous formulas under the form

$$\exists x\ \forall y\ (y < g(x,0)) \wedge (z < 0) \qquad \forall x\ y = x$$

In a first-order formula, some variables occur after a quantifier (existential or universal): the occurrences of these variables are said to be *bounded* and the other occurrences are said to be *free*. For example, in the formula

$$\exists x\ (\underline{y} < h(\underline{x},0)) \wedge \forall y\ (\underline{\underline{z}} < \underline{y})$$

the simply underlined occurrences of x and y are bounded and the occurrences of z and y doubly underlined are free. A variable is free if at least one of its occurrences is free. The set $FV(\varphi)$ of free variables of a formula φ can be defined inductively as follows:

(1) If φ is an atomic formula, $FV(\varphi)$ is the set of variables occurring in φ,
(2) $FV(\neg\varphi) = FV(\varphi)$
(3) $FV(\bigwedge_{i\in I} \varphi_i) = FV(\bigvee_{i\in I} \varphi_i) = \bigcup_{i\in I} FV(\varphi_i)$
(4) $FV(\varphi \to \psi) = FV(\varphi) \cup FV(\psi)$
(5) $FV(\exists x\varphi) = FV(\forall x\varphi) = FV(\varphi) - \{x\}$

A *sentence* is a formula in which all occurrences of variables are bounded. For example, the formula

$$\exists x\ \forall y\ (y < f(x,0))$$

is a sentence.

We let $\varphi(x_1, x_2, \ldots, x_n)$ denote a formula φ in which the set of free variables is contained in $\{x_1, \ldots, x_n\}$ (but is not necessarily equal to $\{x_1, \ldots, x_n\}$).

The variables used in first-order logic, or *first-order* variables, are interpreted, as we shall see, as the elements of a set. In *second-order logic*, one makes use of another type of variables, called *second-order variables*, which represent relations. These variables are traditionally denoted by capital letters: X_0, X_1, etc.. One builds in this way the set of second-order formulas on $\mathcal{L}$, denoted by $\mathbf{SO}[\mathcal{L}]$. The set of terms is the same as for first-order logic. The *atomic formulas* are either of the form

$$(t_1 = t_2)$$

where t_1 and t_2 are terms, or of the form

$$R(t_1, \ldots, t_n) \qquad \text{or} \qquad X(t_1, \ldots, t_n)$$

where $t_1, \ldots, t_n$ are terms, R is an n-ary relation symbol of $\mathcal{L}$ and X is a variable representing a n-ary relation.

Finally, *second-order formulas* on $\mathcal{L}$ form the least set of expressions containing the atomic formulas and closed under the following formation rules:

(i) If φ and ψ are second-order formulas, then so are

$$\neg\varphi, \quad (\varphi \wedge \psi), \quad (\varphi \vee \psi), \quad (\varphi \to \psi)$$

(ii) If φ is a second-order formula, if x is a first-order variable and if X is a relation variable, then the expressions

$$(\exists x\varphi) \qquad (\forall x\varphi) \qquad (\exists X\varphi) \qquad (\forall X\varphi)$$

are second-order formulas.

Monadic second-order logic is the fragment of second-order logic in which the only relation variables are unary relation variables, in other words, variables representing subsets of the domain. By convenience, they are called *set variables*. We let $\mathbf{MSO}(\mathcal{L})$ denote the set of monadic second-order formulas on $\mathcal{L}$. We shall also use the notation $x \in X$ instead of $X(x)$.

2.2 Semantics

We have adopted so far a syntactic point of view to define formulas with no reference to semantics. But of course, formulas would be uninteresting if they were meaningless. Giving a precise meaning to formulas requires to specify the domain on which we want to interpret each of the symbols of the language $\mathcal{L}$. Formally, a *structure* $\mathcal{S}$ on $\mathcal{L}$ is given by a nonempty set D, called *domain* and by a map defined on $\mathcal{L}$, called an *assignment* which associates

(1) with each n-ary relation symbol of $\mathcal{L}$, a n-ary relation defined on D,

(2) with each n-ary function symbol f of $\mathcal{L}$, a n-ary function defined on D,

(3) with each constant symbol c of $\mathcal{L}$, an element of D.

To improve notation, we shall use the same notation for the relation [function, constant] symbols and for the relations [functions, constants] represented by these symbols. The context will allow us to determine easily what the notation stands for. For example, we shall always employ the symbol $<$ to designate the usual order relation on a set of integers, independently of the domain ($\mathbb{N}$, $\mathbb{Z}$, or a subset of $\mathbb{N}$).

We still have to interpret variables. Let us start by first-order variables. Given a fixed structure $\mathcal{S}$ on $\mathcal{L}$, with domain D, a *valuation* on $\mathcal{S}$ is a map ν from the set of variables to the set D. It is then easy to extend ν to a function of the set of terms of $\mathcal{L}$ to D, by induction on the formation rules of terms:

(1) If c is a constant symbol, we put $\nu(c) = c$,

(2) if f is a n-ary function symbol and if $t_1, \ldots, t_n$ are terms,

$$\nu\big(f(t_1, \ldots, t_n)\big) = f(\nu(t_1) \ldots \nu(t_n))$$

If ν is a valuation and a an element of D, we let $\nu\binom{a}{x}$ denote the valuation ν' defined by

$$\nu'(y) = \begin{cases} \nu(y) & \text{if } y \neq x \\ a & \text{if } y = x \end{cases}$$

The notion of interpretation can now be easily formalised. Define, for each first-order formula φ and for each valuation ν, the expressions "the valuation ν satisfies φ in $\mathcal{S}$", or "$\mathcal{S}$ satisfies $\varphi[\nu]$", denoted by $\mathcal{S} \models \varphi[\nu]$, as follows:

(1) $\mathcal{S} \models (t_1 = t_2)[\nu]$ if and only if $\nu(t_1) = \nu(t_2)$

(2) $\mathcal{S} \models R(t_1, \ldots, t_n)[\nu]$ if and only if $\big(\nu(t_1), \ldots, \nu(t_n)\big) \in R$

(3) $\mathcal{S} \models \neg\varphi[\nu]$ if and only if not $\mathcal{S} \models \varphi[\nu]$

(4) $\mathcal{S} \models (\bigwedge_{i \in I} \varphi)[\nu]$ if and only if for each $i \in I$, $\mathcal{S} \models \varphi_i[\nu]$

(5) $\mathcal{S} \models (\bigvee_{i \in I} \varphi)[\nu]$ if and only if there exists $i \in I$, $\mathcal{S} \models \varphi_i[\nu]$

(6) $\mathcal{S} \models (\varphi \to \psi)[\nu]$ if and only if $\mathcal{S} \not\models \varphi[\nu]$ or $\mathcal{S} \models \psi[\nu]$

(7) $\mathcal{S} \models (\exists x\varphi)[\nu]$ if and only if $\mathcal{S} \models \varphi[\nu\binom{a}{x}]$ for some $a \in D$

(8) $\mathcal{S} \models (\forall x\varphi)[\nu]$ if and only if $\mathcal{S} \models \varphi[\nu\binom{a}{x}]$ for each $a \in D$

Note that, actually, the truth of the expression "the valuation ν satisfies φ in $\mathcal{S}$" only depends on the values taken by the free variables of φ. In particular, if φ is a sentence, the choice of the valuation is irrelevant. Therefore, if φ is a sentence, one says that φ is satisfied by $\mathcal{S}$ (or that $\mathcal{S}$ satisfies φ), and denote by $\mathcal{S} \models \varphi$, if, for each valuation ν, $\mathcal{S} \models \varphi[\nu]$.

Next we move to the interpretation of second-order formulas. Given a structure $\mathcal{S}$ on $\mathcal{L}$, with domain D, a *second-order valuation* on $\mathcal{S}$ is a map ν which associates with each first-order variable an element of D and with each n-ary relation variable a subset of D^n (i.e. a n-ary relation on D).

If ν is a valuation and R a subset of D^n, $\nu\binom{R}{X}$ denotes the valuation ν' defined by

$$\nu'(x) = \nu(x) \text{ if } x \text{ is a first-order variable}$$

$$\nu'(Y) = \begin{cases} \nu(Y) & \text{if } Y \neq X \\ R & \text{if } Y = X \end{cases}$$

The notion of an interpretation, already defined for first-order logic, is supplemented by the following rules:

(9) $\mathcal{S} \models (X(t_1, \ldots, t_n))[\nu]$ if and only if $\big(\nu(t_1), \ldots, \nu(t_n)\big) \in \nu(X)$

(10) $\mathcal{S} \models (\exists X\varphi)[\nu]$ if and only if there exists $R \subseteq D^n$, $\mathcal{S} \models \varphi[\nu\binom{R}{X}]$

(11) $\mathcal{S} \models (\forall X\varphi)[\nu]$ if and only if for each $R \subseteq D^n$, $\mathcal{S} \models \varphi[\nu\binom{R}{X}]$

Weak monadic second-order logic has the same formulas as monadic second-order logic, but the interpretation is even more restricted: only valuations which associate with set variables *finite* subsets of the domain D are considered.

Two formulas φ and ψ are said to be *logically equivalent* if, for each structure $\mathcal{S}$ on $\mathcal{L}$, we have $\mathcal{S} \models \varphi$ if and only if $\mathcal{S} \models \psi$.

It is easy to see that the following formulas are logically equivalent:

$$\begin{array}{lllll}
(1) & \varphi \wedge \psi & \text{and} & \neg(\neg\varphi \vee \neg\psi) \\
(2) & \varphi \to \psi & \text{and} & \neg\varphi \vee \psi \\
(3) & \forall x\varphi & \text{and} & \neg(\exists x\ \neg\varphi) \\
(4) & \varphi \vee \psi & \text{and} & \psi \vee \varphi \\
(5) & \varphi \wedge \psi & \text{and} & \psi \wedge \varphi \\
(6) & \varphi \wedge \mathbf{false} & \text{and} & \mathbf{false} \\
(7) & \varphi \vee \mathbf{false} & \text{and} & \varphi
\end{array}$$

Consequently, up to logical equivalence, we may assume that the formulas are built without the symbols $\wedge$, $\to$ and $\forall$.

Logical equivalence also permits one to give a more structured form to formulas. A formula is said to be in *disjunctive normal form* if it can be written as a disjunction of conjunctions of atomic formulas or of negations of atomic formulas, in other words in the form

$$\bigvee_{i\in I} \bigwedge_{j\in J_i} (\varphi_{ij} \vee \neg\psi_{ij})$$

where I and the J_i are finite sets, and φ_{ij} and ψ_{ij} are atomic formulas. The next result is standard and easily proved.

Proposition 2.1. *Every quantifier free formula is logically equivalent to a quantifier free formula in disjunctive normal form.*

A first-order formula is said to be in *prenex normal form* if it is of the form

$$\psi = Q_1x_1\ Q_2x_2\ \ldots\ Q_nx_n\ \varphi$$

where the Q_i are existential or universal quantifiers ($\exists$ or $\forall$) and φ is quantifier-free. The sequence $Q_1x_1\ Q_2x_2\ \ldots\ Q_nx_n$, which can be considered as a word on the alphabet

$$\{\ \exists x_1, \exists x_2, \ldots, \forall x_1, \forall x_2, \ldots\ \},$$

is called the quantifier prefix of ψ. The interest in formulas in prenex normal form comes from the following result.

Proposition 2.2. *Every first-order formula is logically equivalent to a formula in prenex normal form.*

Proof. It suffices to verify that if the variable x does not occur in the formula ψ then

$$\exists x(\varphi \wedge \psi) \equiv (\exists x\varphi) \wedge \psi$$
$$\exists x(\varphi \vee \psi) \equiv (\exists x\varphi) \vee \psi$$

and the same formulas hold for the quantifier $\forall$. Hence it is possible, by renaming the variables, to move the quantifiers to the outside. □

One can also introduce normal forms and a hierarchy for monadic second-order formulas. Thus, one can show that every monadic second-order formula is logically equivalent to a formula of the form

$$\psi = Q_1X_1\ Q_2X_2\ \ldots\ Q_nX_n\ \varphi$$

where the Q_i are existential or universal quantifiers and φ is a first-order formula.

2.3 Logic on words

The logical language that we shall use now was introduced by Büchi under the name of "sequential calculus". To interpret formulas on words, one considers each word as a map associating a letter with each index. Let $u = a_0a_1 \dots a_{n-1}$, where $a_0, \dots, a_{n-1}$ are letters, be a nonempty word on the alphabet A. The *domain* of u, denoted by $\mathrm{Dom}(u)$ is the set

$$\mathrm{Dom}(u) = \{0, \dots, |u| - 1\}$$

Define for each letter $a \in A$ a unary relation $\mathbf{a}$ on the domain of u by

$$\mathbf{a} = \{i \in \mathrm{Dom}(u) \mid a_i = a\}.$$

Finally, let us associate with each word u the structure

$$\mathcal{M}_u = \big(\mathrm{Dom}(u), (\mathbf{a})_{a\in A}, <, S, \min, \max\big),$$

For example, if $u = abbaab$, then $\mathrm{Dom}(u) = \{0, 1, \dots, 5\}$, $\mathbf{a} = \{0, 3, 4\}$ and $\mathbf{b} = \{1, 2, 5\}$. We shall also consider various other nonlogical symbols, notably $<$, S, min and max, that will be interpreted respectively as follows:

(1) the symbol $<$ will represent the usual order;
(2) the symbol S will represent the successor relation on $\mathrm{Dom}(u)$, defined by $S(x, y)$ if and only if $y = x + 1$.
(3) the symbols min and max will represent the minimum and the maximum of the domain: 0 and $|u| - 1$.

From now on, we shall interpret logical formulas on words, that is, on a structure of the form $\mathcal{M}_u$ as explained above. Let φ be a sentence. A nonempty word u *satisfies* φ if the structure $\mathcal{M}_u$ satisfies φ. This is denoted by φ is the set

$$L(\varphi) = \{u \in A^+ \mid u \text{ satisfies } \varphi\}$$

From now on, all the variables will be interpreted as natural numbers. Therefore, we shall use logical equivalence restricted to interpretations with domain of the form $\{0, \dots, n\}$.

In the sequel, we shall mainly consider two logical languages: the language

$$\mathcal{L}_< = \{<\} \cup \{\mathbf{a} \mid a \in A\}$$

will be called the *language of the linear order* and the language

$$\mathcal{L}_S = \{S\} \cup \{\mathbf{a} \mid a \in A\}$$

will be called the *language of the successor*. The atomic formulas of the language of the linear order are of the form

$$\mathbf{a}(x), \qquad x = y, \qquad x < y$$

and those of the language of the successor are of the form

$$\mathbf{a}(x), \qquad x = y, \qquad S(x, y).$$

We let $\mathbf{FO}[<]$ and $\mathbf{MSO}[<]$ respectively denote the set of first-order and monadic second-order formulas of signature $\{<, (\mathbf{a})_{a\in A}\}$. Similarly, we let $\mathbf{FO}[S]$ and

$\mathbf{MSO}[S]$ denote the same sets of formulas of signature $\{S, (\mathbf{a})_{a\in A}\}$. Inside first-order logic, the $\Sigma_n[<]$ (resp. $\Sigma_n[S]$) hierarchy is based on the number of quantifier alternations.

We shall now start the comparison between the various logical languages we have introduced. First of all, the distinction between the signatures S and $<$ is only relevant for first-order logic in view of the following proposition. A relation $R(x_1, \cdots, x_n)$ on natural numbers is said to be *defined* by a formula $\varphi(x_1, \ldots, x_n)$ if, for each nonempty word u and for each $i_1, \ldots, i_n \in \mathrm{Dom}(u)$, one has $R(i_1, \ldots, i_n)$ if and only if $\varphi(i_1, \ldots, i_n)$ is true.

Proposition 2.3. *The successor relation can be defined in* $\mathbf{FO}[<]$*, and the order relation on integers can be defined in* $\mathbf{MSO}[S]$*.*

Proof. The successor relation can be defined by the formula

$$(i < j) \wedge \forall k \, \Big((i < k) \to ((j = k) \vee (j < k))\Big)$$

which states that $j = i + 1$ if i is smaller than j and there exist no element between i and j. The formula $i < j$ can be expressed in $\mathbf{MSO}[S]$ as follows:

$$\exists X \, \Big[\forall x \forall y \, \Big(\big((x \in X) \wedge S(x, y)\big) \to (y \in X)\Big)\Big] \wedge (j \in X) \wedge (i \notin X)$$

which intuitively means that there exists an interval of $\mathrm{Dom}(u)$ of the form $[k, |u| - 1]$ containing j but not i. □

If x is a variable, it is convenient to write $x+1$ $[x-1]$ to replace a variable y subject to the condition $S(x, y)$ $[S(y, x)]$. However, the reader should be aware of the fact that $x + y$ is not definable in $\mathbf{MSO}[<]$.

We shall also use the symbols $\leqslant$ and $\neq$ with their usual interpretations: $x \leqslant y$ stands for $(x < y) \vee (x = y)$ and $x \neq y$ for $\neg(x = y)$.

The symbols min, max can also be defined in $\mathbf{FO}[S]$ with two alternations of quantifiers:

$$\begin{aligned} \min: &\quad \exists \min \ \forall x \ \neg S(x, \min) \\ \max: &\quad \exists \max \ \forall x \ \neg S(\max, x) \end{aligned}$$

We shall sometimes need a parametrized, or relative, notion of satisfaction for a formula. Let φ be a sentence. Let $u = a_0 \cdots a_{n-1}$ be a nonempty word and let $i, j \in \mathrm{Dom}(u)$. Then u is said to *satisfy the formula* φ *between* i *and* j i if $a_i \ldots a_{j-1} \models \varphi$.

Proposition 2.4. *For each sentence* φ*, there exists a formula* $\varphi(x, y)$ *with the same signature, the same order (and, in the case of a first-order formula, the same level in the hierarchy* Σ_n*), having* x *and* y *as unique free variables and which satisfies the following property: for each word* u *and for each* $s, t \in \mathrm{Dom}(u)$*,* $u \models \varphi(s, t)$ *if and only if* u *satisfies* φ *between* s *and* t*.*

Proof. The formulas $\varphi(x, y)$ are built by induction on the formation rules as follows: if φ is an atomic formula, we set $\varphi(x, y) = \varphi$. Otherwise, we set

$$(\neg\varphi)(x, y) = \neg\varphi(x, y)$$

$$(\varphi \vee \psi)(x,y) = \varphi(x,y) \vee \psi(x,y)$$
$$(\exists z\varphi)(x,y) = \exists z\ ((x \leqslant z) \wedge (z < y) \wedge \varphi(x,y))$$
$$(\exists X\varphi)(x,y) = \exists X\ ((\forall z\ (X(z) \to (x \leqslant z) \wedge (z < y)) \wedge \varphi(x,y))$$

It is easy to verify that the formulas $\varphi(x,y)$ built in this way have the required properties. □

3 Monadic second-order logic on words

This section is devoted to the proof of a result of Büchi stating that the subsets of A^* definable in monadic second-order logic are exactly the rational sets.

Theorem 3.5. *A language is definable by a formula of* $\mathbf{MSO}[S]$ *if and only if it recognisable.*

The proof of this result can be decomposed into two parts: passing from automata to formulas, and from formulas to automata. To pass from words to formulas, we simulate the behaviour of an automaton by a formula.

Proposition 3.6. *For each automaton* $\mathcal{A} = (Q, A, E, I, F)$*, there exists a formula* φ *of* $\mathbf{MSO}[<]$ *such that* $L(\varphi) = L^+(\mathcal{A})$*.*

Proof. Suppose that $Q = \{1, \ldots, n\}$. We first write a formula ψ expressing the existence of a path with label u. To this purpose, we associate with each state q a set variable X_q which encodes the set of positions in which a given path visits the state q. The formula states that the X_q's are pairwise disjoint and that if a path is in state q in position x, in state q' in position $x+1$ and if the x-th letter is an a, then $(q,a,q') \in E$. This gives

$$\psi = \Big(\bigwedge_{q \neq q'} \neg\exists x(X_q(x) \wedge X_{q'}(x))\Big) \wedge \Big(\forall x \forall y\ S(x,y) \to \bigvee_{(q,a,q') \in E} \big(X_q(x) \wedge \mathbf{a}(x) \wedge X_{q'}(y)\big)\Big)$$

It remains to state that the path is successful. It suffices to know that min belongs to one of the X_q's such that $q \in I$ and that max belongs to one of the X_q's such that $q \in F$. Therefore, we set

$$\psi_+ = \psi \wedge \Big(\bigvee_{q \in I} X_q(\min)\Big) \wedge \Big(\bigvee_{q \in F} X_q(\max)\Big)$$

The formula

$$\varphi = \exists X_1 \exists X_2 \cdots \exists X_n\ \psi_+$$

now entirely encodes the automaton. □

To pass from sentences to automata, a natural idea is to argue by induction on the formation rules of formulas. The problem is that the set $L(\varphi)$ is only defined when φ is a sentence. The traditional solution in this case consists of adding constants to interpret free variables to the structure in which the formulas are interpreted. For the sake of homogeneity, we proceed in a slightly different way, so that these structures remain words.

The idea is to use an extended alphabet of the form

$$B_{p,q} = A \times \{0,1\}^p \times \{0,1\}^q$$

such that p [q] is greater than or equal to the number of first-order [second-order] variables of φ. A word on the alphabet $B_{p,q}$ can be identified with the sequence

$$(u_0, u_1, \ldots, u_p, u_{p+1}, \ldots, u_{p+q})$$

where $u_0 \in A^*$ and $u_1, \ldots, u_p, u_{p+1}, \ldots, u_{p+q} \in \{0,1\}^*$. We are actually interested in the set $K_{p,q}$ of words of $B^*_{p,q}$ in which each of the components $u_1, \ldots, u_p$ contains exactly one occurrence of 1. If the context permits, we shall write B instead of $B_{p,q}$ and K instead of $K_{p,q}$.

For example, if $A = \{a, b\}$, a word of $B^*_{3,2}$ is represented in Figure 3.1.

u_0	a	b	a	a	b	a	b
u_1	0	1	0	0	0	0	0
u_2	0	0	0	0	1	0	0
u_3	1	0	0	0	0	0	0
u_4	0	1	1	0	0	1	1
u_5	1	1	0	1	0	1	0

Figure 3.1. A word of $B^*_{3,2}$.

The elements of K are called *marked words* on A. This terminology expresses the fact that the elements of K are (finite or infinite) words in which labels marking certain positions have been added. Each of the p first rows only marks one position and the last q ones an arbitrary number of positions.

We first prove the following property.

Proposition 3.7. *For each $p, q \geqslant 0$, the set $K_{p,q}$ is a star-free language of $B^*_{p,q}$.*

Proof. Set, for $1 \leqslant i \leqslant p$,

$$C_i = \{\ (b_0, b_1, \ldots, b_{p+q}) \in B \mid b_i = 1\ \}$$

Then K is the language of B^* containing exactly one letter of each C_i, for $1 \leqslant i \leqslant p$. Now the formula

$$K = \bigcap_{1 \leqslant i \leqslant p} B^* C_i B^* - \bigcup_{1 \leqslant i \leqslant p} B^* C_i B^* C_i B^*.$$

shows that K is a star-free subset of B^*. □

The interpretation of formulas on the words of $B^*_{p,q}$ follows the main lines of the interpretation described in Section 2.3 although the interpretation of $\mathbf{a}$ is slightly modified. Let $u_0 = a_0 a_1 \ldots a_{n-1}$, where $a_0, \ldots, a_{n-1}$ are letters. Then

$$\mathbf{a} = \{i \in \mathrm{Dom}(u_0) \mid a_i = a\}.$$

Let $\varphi(x_1, \ldots, x_r, X_1, \ldots, X_s)$ be a formula in which the first-order [second-order] free variables are $x_1, \ldots, x_r$ [$X_1, \ldots, X_s$], with $r \leqslant p$ and $s \leqslant q$. Let $u =$

$(u_0, u_1, \ldots, u_{p+q})$ be a word of $K_{p,q}$ and, for $1 \leqslant i \leqslant p$, let n_i denote the position of the unique 1 of the word u_i. In the example above, one would have $n_1 = 1$, $n_2 = 4$ and $n_3 = 0$. A word u is said to satisfy φ if u_0 satisfies $\varphi[\nu]$, where ν is the valuation defined by

$$\begin{aligned} \nu(x_j) &= n_j \qquad (1 \leqslant j \leqslant r) \\ \nu(X_j) &= \{\ i \in \mathrm{Dom}(u_0) \mid u_{p+j,i} = 1\ \} \qquad (1 \leqslant j \leqslant s) \end{aligned}$$

In other words, each X_j is interpreted as the set of positions of 1's in u_{p+j}, and each x_j as the unique position of 1 in u_j. Note that for $p = q = 0$, we recover the customary interpretation of sentences.

Set

$$S_{p,q}(\varphi) = \{\ u \in K_{p,q} \mid u \text{ satisfies } \varphi(x_1, \ldots, x_p, X_1, \ldots, X_q)\ \}$$

Again, we shall sometimes simply use the notation $L(\varphi)$ instead of $S_{p,q}(\varphi)$. Conjunctions, disjunctions and negations are easily converted to Boolean operations.

Proposition 3.8. *For each finite family of formulas $(\varphi_i)_{i\in I}$, the following equalities hold:*

(1) $L(\bigvee_{i\in I} \varphi_i) = \bigcup_{i\in I} L(\varphi_i)$,
(2) $L(\bigwedge_{i\in I} \varphi_i) = \bigcap_{i\in I} L(\varphi_i)$,
(3) $L(\neg\varphi) = K_{p,q} - L(\varphi)$

Proof. This is an immediate consequence of the definitions. □

To conclude the proof of Theorem 3.5, it remains to prove by induction on the formation rules of formulas that the sets $L(\varphi)$ are rational. Let us start with the atomic formulas. As for the set K, we prove a slightly stronger result that will be used later on in this chapter.

Proposition 3.9. *For each variable x, y, for each set variable X and for each letter $a \in A$, the sets of the form $L(\mathbf{a}(x))$, $L(x = y)$, $L(x < y)$ and $L(X(x))$ are star-free, and hence rational, subsets of B^*.*

Proof. Set, for $i, j \in \{1, \ldots, p+q\}$

$$\begin{aligned} C_{j,a} &= \{b \in B_{p,q} \mid b_j = 1 \text{ and } b_0 = a\} \\ C_{i,j} &= \{b \in B_{p,q} \mid b_i = b_j = 1\} \\ C_i &= \{b \in B_{p,q} \mid b_i = 1\} \end{aligned}$$

Setting $B = B_{p,q}$, we then have

$$\begin{aligned} L(\mathbf{a}(x_i)) &= K \cap B^* C_{i,a} B^* \\ L(x_i = x_j) &= K \cap B^* C_{i,j} B^* \\ L(x_i < x_j) &= K \cap B^* C_i B^* C_j B^* \\ L(X_i(x_j)) &= K \cap B^* C_{i+p,j} B^* \end{aligned}$$

which establishes the proposition. □

Proposition 3.8 allows one to treat logical connectives. It remains to treat the case of the formulas of the form $\exists x\varphi$ and $\exists X\varphi$. Denote by π_i the function that erases the i-th component, defined by

$$\pi_i(b_0, b_1, \ldots, b_{p+q}) = (b_0, b_1, \ldots, b_{i-1}, b_{i+1}, \ldots, b_{p+q})$$

Thus π_i should be considered as a function from $B_{p,q}$ to $B_{p-1,q}$ if $i \leqslant p$ and to $B_{p,q-1}$ if $p < i \leqslant p+q$.

Proposition 3.10. *For each formula φ, the following formulas hold*

(1) $S_{p-1,q}(\exists x_p\varphi) = \pi_p(S_{p,q}(\varphi))$
(2) $S_{p,q-1}(\exists X_q\varphi) = \pi_{p+q}(S_{p,q}(\varphi))$

Proof. This follows from the definition of existential quantifiers. □

We are now ready to show that if $L(\varphi)$ is rational, then so are $L(\exists x\varphi)$ and $L(\exists X\varphi)$. We may assume that $x = x_p$ and $X = X_q$. Then, by Proposition 3.10, we have $L(\exists x_p\varphi) = \pi_p(L(\varphi))$ and $L(\exists X_q\varphi) = \pi_{p+q}(L(\varphi))$. Since morphisms preserve rationality, the result follows.

This concludes the proof of Büchi's theorem. Note that the proof of Proposition 3.6 shows that the hierarchy on formulas of $\mathbf{MSO}[<]$ based on the number of alternations of second-order quantifiers collapses to the first level. We have indeed the following result.

Proposition 3.11. *Every formula of $\mathbf{MSO}[<]$ is equivalent to a formula of the form*

$$\exists X_1 \ldots \exists X_k \ \varphi(X_1, \ldots, X_k)$$

where φ is a first-order formula.

In fact, one can even show that every formula of $\mathbf{MSO}[<]$ is equivalent to a formula of the form $\exists X\varphi(X)$ where φ is a first-order formula.

4 First-order logic of the linear order

We now study the language $\mathbf{FO}[<]$ of the first-order logic of the linear order. We shall, in a first step, characterise the sets of words definable in this logic, which happen to be the star-free sets. Next, we shall see in Section XX.2 how this result can be refined to establish a correspondence between the levels of the Σ_n-hierarchy of first-order logic and the concatenation hierarchy of star-free sets.

4.1 First order and star-free sets

We shall prove the following key result.

Theorem 4.12. *A language is star-free if and only if it is definable by a formula of $\mathbf{FO}[<]$.*

The first part of the proof consists in converting languages to formulas. To start with, the equalities

$$L(\mathbf{true}) = A^* \qquad L(\mathbf{false}) = \emptyset$$

show that the basic star-free sets are definable by a formula of $\mathbf{FO}[<]$. The Boolean operations are easily converted to connectives as in Proposition 3.8.

We now treat the marked product. We shall use Proposition 2.4 to replace each formula $\varphi \in \mathbf{FO}[<]$ by a formula $\varphi(x, y) \in \mathbf{FO}[<]$ such that for each word $u = a_0 \cdots a_{n-1}$ and each $s, t \in \mathrm{Dom}(u)$, we have $u \models \varphi(s, t)$ if and only if $a_s a_{s+1} \ldots a_t \models \varphi$. The next result, whose proof is immediate, shows that if $X_1 \subset A^*$ and $X_2 \subset A^*$ are definable in $\mathbf{FO}[<]$, then so is the marked product $X_1 a X_2$ for each $a \in A$.

Proposition 4.13. *Let $X_1 \subseteq A^*$ and $X_2 \subseteq A^*$. If $X_1 = L(\varphi_1)$ and $X_2 = L(\varphi_2)$, then $X_1 a X_2 = L(\varphi)$ with $\varphi = \exists y\ \varphi_1(\min, y-1) \wedge \mathbf{a}(y) \wedge \varphi_2(y, \max)$.*

For the opposite implication, from formulas to languages, we show by induction on the formation rules of formulas that $L(\varphi)$ is star-free for every formula $\varphi \in \mathbf{FO}[<]$. In order to treat formulas with free variables, we shall work with the extended alphabets $B_{p,q}$ introduced in Section 3. But since there is no set variables any more, we may assume that $q = 0$, which permits one to eliminate the q indices in the notation. Therefore, we simply set $B_p = A \times \{0, 1\}^p$ and we let K_p (or simply K) denote the set of the words u of B_p^* in which the components $u_1, \ldots, u_p$ contain exactly one occurrence of 1. Note that $K_0 = A^*$, but that K_p does not contain the empty word for $p > 0$.

We also set, for each formula $\varphi = \varphi(x_1, \ldots, x_p)$,

$$S_p(\varphi) = \{u \in K_p \mid u \models \varphi\}$$

By Proposition 3.9, the sets $L(\mathbf{a}(x))$, $L(x = y)$ and $L(x < y)$ are star-free. By Proposition 3.7, the set K is also star-free. The Boolean operations are translated in the usual way by connectives. It remains to treat the case of quantifiers. We shall use the following result.

Proposition 4.14. *Let B be an alphabet and let $B = C \cup D$ be a partition of B. Let X be a star-free language of B^* such that every word of X has exactly one letter in C. Then X can be written in the form*

$$X = \bigcup_{1 \leqslant i \leqslant n} Y_i c_i Z_i$$

with $n \geqslant 0$, $c_i \in C$, and where the $Y_i \subseteq D^*$ and $Z_i \subseteq D^*$ are star-free sets.

Proof. Let $\mu : B^* \to M$ be the syntactic morphism of X. Since X is star-free, Theorem VI.3.3 shows that M is aperiodic. Then

$$X = \bigcup \mu^{-1}(s) c \mu^{-1}(u)$$

where the union runs over all the triples (s, c, u) such that $s \in M$, $c \in C$, $u \in M$ and $s\mu(c)u \in \mu(X)$. Since M is an aperiodic monoid, the sets $\mu^{-1}(s)$ and $\mu^{-1}(u)$ are star-free by Theorem VI.3.3. □

We now treat the case of the existential quantifier. Suppose that $X = L(\varphi)$ is star-free. Denote by $\pi : B_p \to B_{p-1}$ the projection which erases the component corresponding to x_p (we may assume $x = x_p$). Let us apply Proposition 4.14 with $X = L(\varphi)$, $B = B_p$, $C = \{b \in B_p \mid b_p = 1\}$ and $D = \{b \in B_p \mid b_p = 0\}$. The fact that each word of X contains exactly one occurrence in C follows from the inclusion $X \subseteq K$. Therefore

$$\pi(X) = \bigcup_{1 \leqslant i \leqslant n} \pi(Y_i)\pi(b_i)\pi(Z_i)$$

Since the restriction of π to D^* is an isomorphism, the sets $\pi(Y_i)$ and $\pi(Z_i)$ are all star-free. Therefore $L(\exists x\varphi) = \pi(X)$ is a star-free subset. This concludes the proof of Theorem 4.12. □

The proof of Theorem 4.12 given above makes use of the syntactic characterisation of star-free sets. One can also show directly that $L(\varphi)$ is aperiodic for $\varphi \in \mathbf{FO}[<]$.

We can now prove the result announced before.

Corollary 4.15. *One has* $\mathbf{FO}[<] < \mathbf{MSO}[<]$.

Proof. We have already seen that $\mathbf{FO}[<] \leqslant \mathbf{MSO}[<]$. Let φ be a formula of $\mathbf{MSO}[<]$ such that $L(\varphi)$ is not a star-free set. Then, by Theorem 4.12, there is no formula of $\mathbf{FO}[<]$ equivalent to φ. □

Example 4.1. Let φ be the formula

$$\varphi = \exists X \ \big(X(\min) \wedge \forall x \ (X(x) \leftrightarrow \neg X(x+1)) \wedge \neg X(\max)\big)$$

where $\leftrightarrow$ is the connective designating logical equivalence.

A finite word u satisfies φ if and only if $|u|$ is even. Thus $X = L(\varphi)$ is not star-free and hence φ cannot be equivalent to any formula of $\mathbf{FO}[<]$ or, equivalently, X is not definable in $\mathbf{FO}[<]$.

4.2 Σ_1 formulas and piecewise testable languages

Let Σ_1 be the set of *existential* formulas, that is, formulas of the form

$$\exists x_1 \ \exists x_2 \ \ldots \ \exists x_k \ \varphi$$

where φ is quantifier-free.

5 Exercises

Exercise 1. Let A be an alphabet. A language L of A^* is said to be *stutter-invariant* if, for each letter $a \in A$ and for all $x, y \in A^*$, one has $xay \in L$ if and only if $xaay \in L$. This is equivalent to saying that, for each letter $a \in A$, $a \sim_L a^2$. The aim of this problem is to study the class $\mathcal{C}$ of all languages that are both stutter-invariant and piecewise testable.

A language of the form

$$A^*a_1A^*a_2 \cdots A^*a_kA^*$$

where the a_i are letters such that $a_i \neq a_{i+1}$, for $1 \leqslant i \leqslant k-1$, is said to be *elementary stutter-invariant*.

Prove that any Boolean combination of elementary stutter-invariant languages belongs to $\mathcal{C}$.

[More difficult] Prove that, conversely, each language of $\mathcal{C}$ is a Boolean combination of elementary stutter-invariant languages.

Consider the signature $\{\mathbf{a} \mid a \in A\} \cup \{\leqslant\}$, where the symbol $\leqslant$ is interpreted as the usual order relation on integers and each symbol $\mathbf{a}$ has his usual interpretation. Be aware that this signature differs from the usual signature $\{\mathbf{a} \mid a \in A\} \cup \{<, =\}$.

We are interested in the first order fragments $\Sigma_1[\leqslant]$ and $\mathcal{B}\Sigma_1[\leqslant]$.

Give a $\Sigma_1[\leqslant]$-formula defining the language $A^*aA^*bA^*aA^*$ on the alphabet $A = \{a, b, c\}$).

Prove that each elementary stutter-invariant language can be defined by a $\Sigma_1[\leqslant]$ formula.

Prove that a language can be defined by a $\mathcal{B}\Sigma_1[\leqslant]$-formula if and only if it belongs to $\mathcal{C}$.

6 Notes

Exercise 1 was inspired by Kufleitner and Lauser [76].

Part C

The profinite world

Chapter X

Profinite words

The results presented in this chapter are a good illustration of the following quotation of Marshall Stone [153]: *A cardinal principle of modern mathematical research may be stated as a maxim: "One must always topologize".* Indeed, a much deeper insight into the structure of recognisable languages is made possible by the introduction of topological concepts.

1 Topology

We start with a brief reminder of the elements of topology used in the sequel: open and closed sets, continuous functions, metric spaces, compact spaces, etc. This section is by no means a first course in topology but is simply thought as a reminder.

1.1 General topology

Recall that a *topology* on a set E is a set $\mathcal{T}$ of subsets of E satisfying the following conditions:

(1) The empty set and E are in $\mathcal{T}$,

(2) The union of arbitrary many elements of $\mathcal{T}$ is an element of $\mathcal{T}$,

(3) The intersection of finitely many elements of $\mathcal{T}$ is an element of $\mathcal{T}$.

The elements of $\mathcal{T}$ are called the *open sets*. The complement of an open set is called a *closed set*. A set is *clopen* if it is both open and closed.

The *closure* of a subset X of E, denoted by $\overline{X}$, is the intersection of the closed sets containing X. A subset of E is *dense* if its closure is equal to E.

A *topological space* is a set E together with a topology on E. A topology $\mathcal{T}_2$ on a set is a *refinement* of a topology $\mathcal{T}_1$ on the same set, if each open set for $\mathcal{T}_1$ is also an open set for $\mathcal{T}_2$. One also says that $\mathcal{T}_1$ is *coarser* than $\mathcal{T}_2$ or that $\mathcal{T}_2$ is *stronger* than $\mathcal{T}_1$.

The coarsest topology on E is the *trivial* topology, which reduces to the empty set and E. The strongest topology is the *discrete* topology defined by $\mathcal{T} = \mathcal{P}(E)$.

If F is a subset of a topological space $(E, \mathcal{T})$, then the traces $F \cap X$ for $X \in \mathcal{T}$, define a topology on F, called the *relative topology*.

It is sometimes convenient to give a *basis* for a topology on E. This is a collection $\mathcal{B}$ of subsets of E such that every open set is the union of elements of $\mathcal{B}$. Thus $\mathcal{B}$ is a basis for some topology on E if it satisfies the two following conditions:

(i) E is the union of all the elements of $\mathcal{B}$,

(ii) every finite intersection of elements of $\mathcal{B}$ is a union of elements of $\mathcal{B}$.

The open sets of the topology generated by $\mathcal{B}$ are by definition the arbitrary unions of elements of $\mathcal{B}$.

A map from a topological space to another one is *continuous* if the inverse image of each open set is an open set. It is an *homeomorphism* if it is a continuous bijection and the inverse bijection is also continuous. Two topological spaces are *homeomorphic* if there is an homeomorphism between them.

Let $(E_i, \mathcal{T}_i)_{i\in I}$ be a family of topological spaces, and let $E = \prod_{i\in I} E_i$ be the cartesian product of the E_i's. Denote by π_i the natural projection from E onto E_i, defined by $\pi_i((e_j)_{j\in I}) = e_i$. The *product topology* on E is the topology generated by the basis consisting of the finite intersections of sets of the form $\pi_i^{-1}(X_i)$ where X_i is an open set of E_i. These sets are nothing else than the products $\prod_{i\in I} X_i$ where the X_i's are open sets of E_i and where $X_i = E_i$ except for a finite number of indices. The natural projections $\pi_i : E \to E_i$ are then continuous and a mapping $\varphi : F \to E$ is continuous if and only if the mappings $\pi_i \circ \varphi : F \to E_i$, for $i \in I$, are all continuous.

A topological space $(E, \mathcal{T})$ is *Hausdorff* if for each $u, v \in E$ with $u \neq v$, there exist *disjoint* open sets U and V such that $u \in U$ and $v \in V$. If f is a continuous map from a topological space X to an Hausdorff space Y, then the graph of f is closed in $X \times Y$.

1.2 Metric spaces

A *metric* d on a set E is a map $d : E \times E \to \mathbb{R}_+$ from $E \times E$ to the set of nonnegative real numbers satisfying the following three conditions, for every $x, y, z \in E$:

(1) $d(x, y) = 0$ if and only if $x = y$,

(2) $d(y, x) = d(x, y)$,

(3) $d(x, z) \leqslant d(x, y) + d(y, z)$

A *metric space* is a set E together with a metric d on E.

The topology defined by d is obtained by taking as a basis the *open ε-balls* defined for $x \in E$ and $\varepsilon > 0$ by

$$B(x, \varepsilon) = \{y \in E \mid d(x, y) < \varepsilon\}$$

Every metric space is Hausdorff. Indeed, given two distinct elements x and y, the open balls $B(x, \varepsilon)$ and $B(y, \varepsilon)$ are disjoint if ε is set to be half the distance between x and y.

A sequence $(x_n)_{n\geqslant 0}$ of elements of a metric space (E, d) is converging to a *limit* x if, for each $\varepsilon > 0$, there exists an integer k such that for each $n \geqslant k$, $d(x_n, x) < \varepsilon$. A sequence is said to be *convergent* if it admits a limit.

Limits are convenient when characterising other topological properties of metric spaces. For instance, a subset F of E is closed if and only if, for every sequence $(x_n)_{n\geqslant 0}$ of elements of F converging to a limit x, x itself belongs to

F. A map $\varphi : (E, d) \to (E', d')$ between two metric spaces is continuous if and only if, for every sequence $(x_n)_{n\geqslant 0}$ of elements of E converging to a limit x, the sequence $(\varphi(x_n))_{n\geqslant 0}$ converges to $\varphi(x)$.

A *Cauchy sequence* in a metric space (E, d) is a sequence $(x_n)_{n\geqslant 0}$ of elements of E such that for each $\varepsilon > 0$, there exists an integer k such that for each $n \geqslant k$ and $m \geqslant k$, $d(x_n, x_m) < \varepsilon$. Every convergent sequence is a Cauchy sequence, but the converse does not hold in general. A metric space in which each Cauchy sequence is convergent is said to be *complete*. Every closed subset of a complete space is complete.

Let $(E_i, d_i)_{1\leqslant i\leqslant n}$ be a finite family of metric spaces. Then $(E_1 \times \cdots \times E_n, d)$ is a metric space, where d, defined by

$$d((x_1, \ldots, x_n), (y_1, \ldots, y_n)) = \max\{d(x_1, y_1), \ldots, d(x_n, y_n)\}$$

is a metric that defines the product topology.

Let (E, d) and (E', d') be two metric spaces. A function φ from E to E' is said to be *uniformly continuous* if for each $\varepsilon > 0$, there exists $\delta > 0$ such that the relation $d(x, y) < \delta$ implies $d'(\varphi(x), \varphi(y)) < \varepsilon$. If φ is uniformly continuous, the image under φ of a Cauchy sequence of E is a Cauchy sequence of E'. We say that φ is a *uniform homeomorphism* if it is a uniformly continuous bijection and φ^{-1} is also uniformly continuous. Two metric spaces are *uniformly homeomorphic* if there is a uniform homeomorphism between them.

We say that φ is an *isometry* if it is a bijection from E onto E' such that, for each $x, y \in E$,

$$d(x, y) = d'(\varphi(x), \varphi(y))$$

The *completion* of a metric space E is a complete metric space $\widehat{E}$ together with an isometric embedding of E as a dense subspace of $\widehat{E}$. One can prove that every metric space admits a completion, which is unique up to isometric equivalence: if $\widehat{E}_1$ and $\widehat{E}_2$ are two completions of E, there exists an isometry of $\widehat{E}_1$ onto $\widehat{E}_2$, whose restriction to E is the identity.

The completion of E can be constructed as follows. Let $C(E)$ be the set of Cauchy sequences in E. Define an equivalence relation $\sim$ on $C(E)$ as follows. Two Cauchy sequences $x = (x_n)_{n\geqslant 0}$ and $y = (y_n)_{n\geqslant 0}$ are equivalent if the interleaved sequence $x_0, y_0, x_1, y_1, \ldots$ is also a Cauchy sequence. The completion of E is defined to be the set $\widehat{E}$ of equivalence classes of $C(E)$. The metric d on E extends to a metric on $\widehat{E}$ defined by

$$d(x, y) = \lim_{n\to\infty} d(x_n, y_n)$$

where x and y are representative Cauchy sequences of elements in $\widehat{E}$. The definition of the equivalence ensures that the above definition does not depend on the choice of x and y in their equivalence class and the fact that $\mathbb{R}$ is complete ensures that the limit exists.

1.3 Compact spaces

An important notion is that of *compact* space. A family of open sets $(U_i)_{i\in I}$ is said to *cover* a topological space $(E, \mathcal{T})$ if $E = \bigcup_{i\in I} U_i$. A topological space $(E, \mathcal{T})$ is said to be *compact* if for each family of open sets covering E, there exists a finite subfamily that still covers E.

Compact metric spaces admit two important characterisations. First, a metric space is compact if and only if every sequence has a convergent subsequence. Second, a metric space is compact if and only if it is complete and *totally bounded*, which means that, for every $n > 0$, the space can be covered by a finite number of open balls of radius $< 2^{-n}$. Furthermore, a metric space is totally bounded if and only if its completion is totally bounded. Therefore a metric space is totally bounded if and only if its completion is compact.

If X is a closed set of a compact space E, then X, equipped with the relative topology, is also a compact space. We shall use freely the well-known result that every product of compact spaces is compact (Tychonov's theorem).

One can show that if $\varphi : E \to F$ is a continuous map from a topological space E to a topological Hausdorff space F, the image of a compact set under φ is still compact. Moreover if E is a compact metric space and F is a metric space, then every continuous function from E to F is uniformly continuous.

We conclude this section with a useful extension result that is worth to be stated separately.

Proposition 1.1. *Let E and F be metric spaces. Any uniformly continuous function $\varphi : E \to F$ admits a unique uniformly continuous extension $\widehat{\varphi} : \widehat{E} \to \widehat{F}$. Furthermore, if $\widehat{E}$ is compact and if φ is surjective, then $\widehat{\varphi}$ is surjective.*

In particular, if F is complete, any uniformly continuous function from E to F admits a unique uniformly continuous extension from $\widehat{E}$ to F.

1.4 Topological semigroups

A *topological semigroup* is a semigroup S equipped with a topology for which the semigroup operation is continuous. This means that the function from $S \times S$ to S which maps (x, y) onto xy is continuous. A *compact semigroup* is a topological semigroup which is compact as a topological space.

2 Profinite topology

2.1 The profinite metric

Let A be a finite alphabet. A morphism $\varphi : A^* \to M$ *separates* two words u and v of A^* if $\varphi(u) \neq \varphi(v)$. By extension, we say that a monoid M *separates* two words if there is a morphism from A^* onto M that separates them.

Example 2.1.

(1) The words $ababa$ and $abaa$ can be separated by a group of order 2. Indeed, let $\pi : A^* \to \mathbb{Z}/2\mathbb{Z}$ be the morphism defined by $\pi(x) = |x| \pmod 2$. Then $\pi(ababa) = 1$ and $\pi(abaa) = 0$ and hence π separates u and v.

(2) More generally, two words u and v of unequal length can be separated by a finite cyclic group. Indeed, suppose that $|u| < |v|$ and let $n = |v|$. Let $\pi : A^* \to \mathbb{Z}/n\mathbb{Z}$ be the morphism defined by $\pi(x) = |x| \pmod n$. Then $\pi(v) = 0$ but $\pi(u) \neq 0$. Note that u and v can also be separated by a finite monoid of size $|u|+2$. Define an addition $\oplus$ on $M = \{0, 1, \ldots, |u|+1\}$ by $s \oplus t = \min\{s + t, |u| + 1\}$ and let $\varphi : A^* \to M$ be the morphism defined by $\varphi(x) = \min\{|x|, |u| + 1\}$. Then $\varphi(u) = |u|$ and $\varphi(v) = |u| + 1$.

(3) A similar idea can be applied if the number of occurrences of some letter a is not the same in u and v. Assume for instance that $|u|_a < |v|_a$ and let $n = |v|_a$. Then the morphism $\pi : A^* \to \mathbb{Z}/n\mathbb{Z}$ defined by $\pi(a) = 1$ and $\pi(c) = 0$ for $c \neq a$ separates u and v.

(4) Recall (see Section II.2.2) that the monoid U_2 is defined on the set $\{1, a, b\}$ by the operation $aa = ba = a$, $bb = ab = b$ and $1x = x1 = x$ for all $x \in \{1, a, b\}$. Let u and v be words of $\{a, b\}^*$. Then the words ua and vb can be separated by the morphism $\pi : A^* \to U_2$ defined by $\pi(a) = a$ and $\pi(b) = b$ since $\pi(ua) = a$ and $\pi(ub) = b$.

These examples are a particular case of a general result.

Proposition 2.2. *Any pair of distinct words of A^* can be separated by a finite monoid.*

Proof. Let u and v be two distinct words of A^*. Since the language $\{u\}$ is recognisable, there exists a morphism φ from A^* onto a finite monoid M which recognises it, that is, such that $\varphi^{-1}(\varphi(u)) = \{u\}$. It follows that $\varphi(v) \neq \varphi(u)$ and thus φ separates u and v. □

We now define a metric on A^* with the following intuitive idea in mind: two words are close for this metric if a large monoid is required to separate them. Let us now give the formal definition. Given two words $u, v \in A^*$, we set

$$r(u, v) = \min \{|M| \mid M \text{ is a monoid that separates } u \text{ and } v\}$$
$$d(u, v) = 2^{-r(u,v)}$$

with the usual conventions $\min \emptyset = +\infty$ and $2^{-\infty} = 0$.

Thus $d(u, v) < 2^{-n}$ if and only if u and v cannot be separated by a monoid of size $\leqslant n$ and $d(u, v) < \varepsilon$ if and only if u and v cannot be separated by a monoid of size $\leqslant n$, where n is the unique positive integer such that $2^{-(n+1)} < \varepsilon \leqslant 2^{-n}$.

The following proposition establishes the main properties of d.

Proposition 2.3. *The function d is an ultrametric, that is, satisfies the following properties, for all $u, v, w \in A^*$,*

(1) $d(u, v) = 0$ *if and only if* $u = v$,

(2) $d(u, v) = d(v, u)$,

(3) $d(u, w) \leqslant \max\{d(u, v), d(v, w)\}$.

It also satisfies the property

(4) $d(uw, vw) \leqslant d(u, v)$ *and* $d(wu, wv) \leqslant d(u, v)$.

Proof. (1) follows from Proposition 2.2.

(2) is trivial.

(3) Let M be a finite monoid separating u and w. Then M separates either u and v, or v and w. It follows that $\min(r(u, v), r(v, w)) \leqslant r(u, w)$ and hence $d(u, w) \leqslant \max\{d(u, v), d(v, w)\}$.

(4) A finite monoid separating uw and vw certainly separates u and v. Therefore $d(uw, vw) \leqslant d(u, v)$ and, dually, $d(wu, wv) \leqslant d(u, v)$. □

Thus (A^*, d) is a metric space, but it is not very interesting as a topological space.

Proposition 2.4. *The topology defined on A^* by d is discrete: every subset is clopen.*

Proof. Let u be a word of A^* and let n be the size of the syntactic monoid M of the language $\{u\}$. Then if $d(u,v) < 2^{-n}$, $r(u,v) > n$ and in particular, M does not separate u and v. It follows that $u = v$. Therefore the open ball $B(u, 2^{-n})$ is equal to $\{u\}$. It follows that every singleton is open. Now if U is a language of A^*, one has $U = \bigcup_{u\in U}\{u\}$ and hence U is open. For the same reason, U^c is open and thus every subset of A^* is clopen. □

However, the open balls (A^*, d) have an interesting property. For each $n > 1$, consider the congruence $\sim_n$ defined on A^* by

$$u \sim_n v \text{ if and only if } \varphi(u) = \varphi(v) \text{ for every morphism } \varphi \text{ from } A^* \text{ onto a monoid of size } \leqslant n.$$

Since A is finite, there are only finitely many morphisms from A^* onto a monoid of size $\leqslant n$, and thus $\sim_n$ is a congruence of finite index. Furthermore, $d(u,v) < 2^{-n}$ if and only if u and v cannot be separated by a monoid of size $\leqslant n$, i.e. are $\sim_n$-equivalent.

This simple observation has far-reaching consequences.

Proposition 2.5. *Every open ball of (A^*, d) is a recognisable language and every recognisable language is a finite union of open balls.*

Proof. Let u be a word and let ε be a positive real number. Let n be the unique positive integer such that $2^{-(n+1)} < \varepsilon \leqslant 2^{-n}$ and let $\pi_n : A^* \to A^*/\sim_n$ be the quotient morphism. Then the formula

$$B(u,\varepsilon) = B(u, 2^{-n}) = \{v \in A^* \mid v \sim_n u\} = \pi_n^{-1}(\pi_n(u)) \tag{2.1}$$

shows that $B(u,\varepsilon)$ is a recognisable language.

Conversely, let L be a recognisable language and let $\eta : A^* \to M$ be its syntactic morphism. Let $n = |M|$. Since $u \sim v$ implies $\eta(u) = \eta(v)$ there exists by Proposition II.3.22 a morphism $\gamma : A^*/\sim_n \to M$ such that $\eta = \gamma \circ \pi_n$. Setting $P = \eta(L)$, one gets

$$L = \eta^{-1}(P) = (\gamma \circ \pi_n)^{-1}(P) = \pi_n^{-1}(\gamma^{-1}(P)) = \bigcup_{m\in\gamma^{-1}(P)} \pi_n^{-1}(m)$$

Now, if $\pi_n(u) = m$, then $\pi_n^{-1}(m) = B(u, 2^{-n})$ and thus L is a finite union of open balls. □

Proposition 2.6. *The* metric space *(A^*, d) is totally bounded: for every $n > 0$, A^* is covered by a finite number of open balls of radius $< 2^{-n}$.*

Proof. Indeed the $\sim_n$-classes are open balls of radius $< 2^{-n}$ and cover A^*. □

Here is an important consequence of Proposition 2.6.

Theorem 2.7. *A map $\varphi : A^* \to B^*$ is uniformly continuous if and only if, for every recognisable language of B^*, the language $\varphi^{-1}(L)$ is also recognisable.*

Proof. Let $\varphi : A^* \to B^*$ be uniformly continuous. Let us show that if L is a recognisable language of B^*, then $\varphi^{-1}(L)$ is also recognisable. By Proposition 2.6, every recognisable language is a finite union of open balls, and hence it suffices to prove the result when L is an open ball, say $L = B(x, \varepsilon)$.

Since φ is uniformly continuous, there exists n such that if $d(u, v) < 2^{-n}$, then $d(\varphi(u), \varphi(v)) < \varepsilon$. We claim that if $u \in \varphi^{-1}(L)$ and $v \in B(u, 2^{-n})$, then $v \in \varphi^{-1}(L)$. Indeed, since $\varphi(u) \in L$, one has $d(x, \varphi(u)) < \varepsilon$. Furthermore, $d(u, v) < 2^{-n}$ implies that $d(\varphi(u), \varphi(v)) < \varepsilon$. It follows that

$$d(x, \varphi(v)) \leqslant \max\big(d(x, \varphi(u)), d(\varphi(u), \varphi(v))\big) < \varepsilon$$

and thus $\varphi(v) \in L$, which proves the claim. Consequently,

$$\varphi^{-1}(L) \subseteq \bigcup_{u \in \varphi^{-1}(L)} B(u, 2^{-n}) \subseteq \varphi^{-1}(L)$$

It follows by (2.1) that

$$\varphi^{-1}(L) = \bigcup_{u \in \varphi^{-1}(L)} B(u, 2^{-n}) = \bigcup_{u \in \varphi^{-1}(L)} \pi_n^{-1}\big(\pi_n(u)\big) = \pi_n^{-1}\big(\pi_n(\varphi^{-1}(L))\big)$$

and thus $\varphi^{-1}(L)$ is recognisable.

Suppose now that if L is a recognisable language of B^*, then $\varphi^{-1}(L)$ is also recognisable. For each $n > 0$, let $\mathcal{L}_n$ be the (finite) set of all $\sim_n$-classes. If $L \in \mathcal{L}_n$, then L is recognisable, and thus $\varphi^{-1}(L)$ is also recognisable. Let $k(L)$ be the size of its syntactic monoid and let

$$k = \max\{k(L) \mid L \in \mathcal{L}_n\}.$$

We claim that if $d(u, v) < 2^{-k}$, then $d(\varphi(u), \varphi(v)) < 2^{-n}$. Indeed, since $\mathcal{L}_n$ is a partition of B^*, one has $\bigcup_{L \in \mathcal{L}_n} \varphi^{-1}(L) = A^*$. Let L be an element of $\mathcal{L}_n$ such that $u \in \varphi^{-1}(L)$. Since $d(u, v) < 2^{-k}$, the syntactic monoid of $\varphi^{-1}(L)$ does not separate u from v and hence $v \in \varphi^{-1}(L)$. Thus both $\varphi(u)$ and $\varphi(u)$ are in L, whence $u \sim_n v$. It follows that $d(\varphi(u), \varphi(v)) < 2^{-n}$, which proves the claim and shows that φ is uniformly continuous. □

2.2 The free profinite monoid

Let $\widehat{A^*}$ denote the completion of (A^*, d). Its elements are called the *profinite words* on the alphabet A. Since the completion of a totally bounded metric space is compact, Proposition 2.6 gives immediately the following important result.

Theorem 2.8. *The set of profinite words $\widehat{A^*}$ is compact.*

The density of A^* in $\widehat{A^*}$ has several useful consequences, which are summarised in the next propositions. We first establish an important property of the product.

Proposition 2.9. *The function $(u, v) \to uv$ from $A^* \times A^*$ to A^* is uniformly continuous.*

Proof. By Proposition 2.3, one has for all $u, u', v, v' \in A^*$,

$$d(uv, u'v') \leqslant \max\{d(uv, uv'), d(uv', u'v')\} \leqslant \max\{d(v, v'), d(u, u')\}$$

which proves the result. □

It follows from Proposition 1.1 that the product on A^* can be extended in a unique way, by continuity, to $\widehat{A^*}$. Since the formulas $(xy)z = x(yz)$ and $1x = x = x1$ are preserved by passage to the limit, this extended product is also associative and admits the empty word as identity element. The product on $\widehat{A^*}$ is also uniformly continuous and makes $\widehat{A^*}$ a topological monoid. It is called the *free profinite monoid* because it satisfies a universal property comparable to that of A^*. Before stating this universal property precisely, let us state another useful consequence of the density of A^*.

Proposition 2.10. *Every morphism φ from A^* to a discrete finite monoid M is uniformly continuous and can be extended in a unique way to a uniformly continuous morphism $\widehat{\varphi}$ from $\widehat{A^*}$ to M.*

Proof. If $d(u, v) < 2^{-|M|}$, then φ does not separate u and v and hence $\varphi(u) = \varphi(v)$. It follows that φ is a uniformly continuous function from A^* to the discrete metric space M. Therefore by Proposition 1.1, φ has a unique uniformly continuous extension $\widehat{\varphi}$ from $\widehat{A^*}$ to M. It remains to prove that $\widehat{\varphi}$ is a morphism. Let

$$D = \{(u, v) \in \widehat{A^*} \times \widehat{A^*} \mid \widehat{\varphi}(uv) = \widehat{\varphi}(u)\widehat{\varphi}(v)\}$$

We claim that $D = \widehat{A^*} \times \widehat{A^*}$, which exactly says that $\widehat{\varphi}$ is a morphism. We already have $\widehat{\varphi}(uv) = \widehat{\varphi}(u)\widehat{\varphi}(v)$ for $u, v \in A^*$ since φ is a morphism, and thus D contains $A^* \times A^*$. Since $A^* \times A^*$ is dense in $\widehat{A^*} \times \widehat{A^*}$, it suffices to prove that D is closed, which essentially follows from the uniform continuity of the product. In more details, let $\pi : A^* \times A^* \to A^*$ be the map defined by $\pi(u, v) = uv$. Proposition 2.9 shows that π is uniformly continuous and has a unique continuous extension $\widehat{\pi}$ from $\widehat{A^*} \times \widehat{A^*}$ to $\widehat{A^*}$ (the product on $\widehat{A^*}$). With this notation in hand, we get

$$\widehat{\varphi}(uv) = (\widehat{\varphi} \circ \widehat{\pi})(u, v) \text{ and } \widehat{\varphi}(u)\widehat{\varphi}(v) = (\widehat{\pi} \circ (\widehat{\varphi} \times \widehat{\varphi}))(u, v)$$

It follows that

$$D = \bigcup_{m \in M} (\widehat{\varphi} \circ \widehat{\pi})^{-1}(m) \cap (\widehat{\pi} \circ (\widehat{\varphi} \times \widehat{\varphi}))^{-1}(m)$$

Since $\widehat{\varphi} \circ \widehat{\pi}$ and $\widehat{\pi} \circ (\widehat{\varphi} \times \widehat{\varphi})$ are both continuous and $\{m\}$ is a closed subset of M, it follows that D is closed, which concludes the proof. □

However, there are some noncontinuous morphisms from $\widehat{A^*}$ onto a finite monoid. For instance, the morphism $\varphi : \widehat{A^*} \to U_1$ defined by

$$\varphi(u) = \begin{cases} 1 & \text{if } u \in A^* \\ 0 & \text{otherwise} \end{cases}$$

is not continuous since $\varphi^{-1}(1) = A^*$ is not closed. Now, the restriction of φ to A^*, which is continuous, has a continuous extension to $\widehat{A^*}$. But this extension maps every profinite word to 1 and is therefore not equal to φ.

2.3 Universal property of the free profinite monoid

We are ready to state the universal property of $\widehat{A^*}$.

Proposition 2.11. *If φ is a function from A to a finite monoid M, there exists a unique (uniformly) continuous monoid morphism $\widehat{\varphi} : \widehat{A^*} \to M$ such that, for each $a \in A$, $\widehat{\varphi}(a) = \varphi(a)$. Moreover, $\widehat{\varphi}$ is surjective if and only if the set $\varphi(A)$ generates M.*

Proof. First, there exists by Proposition II.5.28 a unique morphism from A^* to M which extends φ and by Proposition 2.10, a unique uniformly continuous morphism $\widehat{\varphi}$ from $\widehat{A^*}$ to M that extends φ.

One has $\widehat{\varphi}(\widehat{A^*}) \subseteq \overline{\varphi(A^*)}$. Since M is discrete, it follows that $\widehat{\varphi}$ is surjective if and only if $\varphi(A)$ generates M. □

An argument similar to the one in the proof of Proposition 2.11 would lead to the following result:

Proposition 2.12. *Every morphism $\varphi : A^* \to B^*$ is uniformly continuous and can be extended in a unique way to a uniformly continuous morphism $\widehat{\varphi} : \widehat{A^*} \to \widehat{B^*}$.*

One can also use Proposition 2.10 to define directly the metric on $\widehat{A^*}$. Let us say that a morphism φ from A^* onto a finite monoid M *separates* two profinite words u and v of $\widehat{A^*}$ if $\widehat{\varphi}(u) \neq \widehat{\varphi}(v)$.
Given two profinite words $u, v \in \widehat{A^*}$, we set

$$r(u,v) = \min\{|M| \mid M \text{ is a finite monoid that separates } u \text{ and } v\}$$
$$d(u,v) = 2^{-r(u,v)}$$

with the usual conventions $\min \emptyset = +\infty$ and $2^{-\infty} = 0$.

The term *profinite* is justified by the following property, which often serves as a definition in the literature.

Proposition 2.13. *The* profinite topology *on $\widehat{A^*}$ is the least topology which makes continuous every morphism from A^* onto a finite discrete monoid.*

Proof. Proposition 2.10 shows that every morphism from A^* onto a finite discrete monoid is continuous for the topology defined by d.

Suppose now that $\widehat{A^*}$ is equipped with a topology $\mathcal{T}$ such that every morphism from A^* onto a finite discrete monoid is continuous. We claim that for all $\varepsilon > 0$ and for all $x \in \widehat{A^*}$, the ball

$$B(x,\varepsilon) = \{y \in \widehat{A^*} \mid d(x,y) < \varepsilon\}$$

is open in $\mathcal{T}$. First observe that $d(x,y) < \varepsilon$ if and only if x and y cannot be separated by a monoid of size $< n$, where n is the smallest positive integer such that $2^{-n} < \varepsilon$. It follows that $B(x,\varepsilon)$ is the intersection of the sets $\varphi^{-1}(\varphi(x))$, where φ runs over the set of all morphisms from A^* onto a monoid of size $< n$. By assumption on $\mathcal{T}$, each set of the form $\varphi^{-1}(\varphi(x))$ is open and since A is finite, there are finitely many morphisms from A^* onto a monoid of size $< n$. This proves the claim and the proposition. □

What about sequences? First, every profinite word is the limit of a Cauchy sequence of words. Next, a sequence of profinite words $(u_n)_{n\geqslant 0}$ is converging to a profinite word u if and only if, for every morphism φ from A^* onto a finite monoid, $\widehat{\varphi}(u_n)$ is ultimately equal to $\widehat{\varphi}(u)$.

2.4 ω-terms

It is relatively difficult to give "concrete" examples of profinite words which are not words. One such example is the profinite word x^ω, associated with every profinite word x. The formal definition is

$$x^\omega = \lim_{n\to\infty} x^{n!}$$

and is justified by the following result, which shows that the sequence $x^{n!}$ has indeed a limit in $\widehat{A^*}$.

Proposition 2.14. *For each $x \in \widehat{A^*}$, the sequence $(x^{n!})_{n\geqslant 0}$ is a Cauchy sequence. It converges to an idempotent element of $\widehat{A^*}$.*

Proof. For the first part of the statement, it suffices to show that for $p, q \geqslant n$, $x^{p!}$ and $x^{q!}$ cannot be separated by a monoid of size $\leqslant n$. Let indeed $\varphi : A^* \to M$ be a monoid morphism, with $|M| \leqslant n$, and put $s = \varphi(x)$. Since M is finite, s has an idempotent power $e = s^r$, with $r \leqslant n$. By the choice of p and q, the integer r divides simultaneously $p!$ and $q!$. Consequently, $s^{p!} = s^{q!} = e$, which shows that M cannot separate $x^{p!}$ and $x^{q!}$.

For n large enough, we also have $\varphi(x^{n!})\varphi(x^{n!}) = ee = e = \varphi(x^{n!})$. It follows that the limit of the sequence $(x^{n!})_{n\geqslant 0}$ is idempotent. □

Note that x^ω is simply a notation and one should resist the temptation to interpret it as an infinite word. The right intuition is to interpret ω as the exponent of a finite semigroup. To see this, let us compute the image of x^ω under a morphism to a finite monoid.

Let M be a finite monoid with exponent ω, let $\varphi : A^* \to M$ a morphism and let $s = \varphi(x)$. Then the sequence $s^{n!}$ is ultimately equal to the unique idempotent s^ω of the subsemigroup of M generated by s. Consequently, we obtain the formula

$$\widehat{\varphi}(x^\omega) = \varphi(x)^\omega$$

which justifies the notation x^ω. Note also that $x^\omega x^\omega = x^\omega$ and $(x^\omega)^\omega = x^\omega$.

Two related examples are the profinite words $x^{\omega+1}$ and $x^{\omega-1}$, which are defined in a natural way by the formulas

$$x^{\omega+1} = \lim_{n\to\infty} x^{n!+1} \quad \text{and} \quad x^{\omega-1} = \lim_{n\to\infty} x^{n!-1}$$

It follows immediately from the definition that $xx^\omega = x^{\omega+1} = x^\omega x$ and that $xx^{\omega-1} = x^\omega = x^{\omega-1}x$. With the notation of the previous paragraph, one also gets $\varphi(x^{\omega+1}) = s^{\omega+1}$ but the interpretation of $x^{\omega-1}$ is a little bit more subtle. Let us first recall the structure of the subsemigroup generated by s: its minimal ideal is a group G whose identity is s^ω. It is tempting to write $\varphi(x^{\omega-1}) = s^{\omega-1}$, but it may happen that $s^{\omega-1}$ is not in G. Worse, if S is idempotent, its exponent is 1 and $s^{\omega-1}$ is not even defined. In fact, in M one has $\lim_{n\to\infty} s^{n!-1} = s^{2\omega-1}$ and thus the right formula is $\varphi(x^{\omega-1}) = s^{2\omega-1}$.

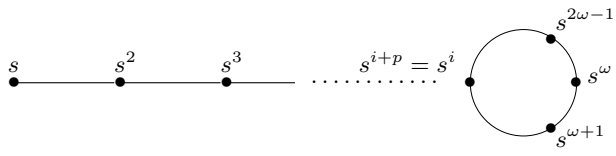


Figure 2.1. The semigroup generated by s.

Note that $s^{2\omega-1}$ is the inverse of $s^{\omega+1}$ in G and this is the right interpretation. Indeed, one can show that in the free profinite monoid, x generates a compact semigroup whose minimal ideal is a monogenic compact group with identity x^ω. Then $x^{\omega-1}$ is the inverse of $x^{\omega+1}$ in this group.

The *ω-terms* on A form the smallest submonoid of $\widehat{A^*}$ containing A^* and closed under the operations $x \to x^\omega$, $x \to x^{\omega+1}$ and $x \to x^{\omega-1}$. For instance, if $A = \{a, b, c\}$, abc, a^ω and $((ab^\omega c)^{\omega-1}ab)^\omega$ are examples of ω-terms. These ω-terms represent the most intuitive examples of profinite words, but unfortunately, there are way more profinite words than ω-terms. To be precise, the free profinite monoid $\widehat{A^*}$ is uncountable (even on a one letter alphabet!) while the set of ω-terms is countable.

3 Recognisable languages and clopen sets

There is a strong connection between recognisable and clopen subsets. We start by considering the recognisable subsets of $\widehat{A^*}$.

Proposition 3.15. *Let P be a subset of $\widehat{A^*}$ and let M be its syntactic monoid. The following conditions are equivalent:*

(1) *P is clopen,*

(2) *the syntactic congruence of P is a clopen subset of $\widehat{A^*} \times \widehat{A^*}$,*

(3) *P is recognisable and its syntactic morphism is a continuous map from $\widehat{A^*}$ onto the discrete finite space M.*

Proof. Let $\sim_P$ denote the syntactic congruence of P and by $\widehat{\eta} : \widehat{A^*} \to M$ its syntactic morphism. Recall that $s \sim_P t$ if, for all $u, v \in \widehat{A^*}$, the conditions $usv \in P$ and $utv \in P$ are equivalent.

(1) implies (2). It follows from the definition of $\sim_P$ that

$$\sim_P = \bigcap_{u,v \in \widehat{A^*}} \left((u^{-1}Pv^{-1} \times u^{-1}Pv^{-1}) \cup (u^{-1}P^cv^{-1} \times u^{-1}P^cv^{-1})\right) \qquad (3.1)$$

If P is clopen, each set $u^{-1}Pv^{-1}$ is also clopen. Indeed, $u^{-1}Pv^{-1}$ is the inverse image of the clopen set P under the continuous function $x \mapsto uxy$. Now, Formula (3.1) shows that $\sim_P$ is closed.

In order to show that the complement of $\sim_P$ is closed, consider a sequence (s_n, t_n) of elements of $(\sim_P)^c$, converging to a limit (s, t). Since $s_n \not\sim_P t_n$, there exist some profinite words u_n, v_n such that $u_n s_n v_n \in P$ and $u_n t_n v_n \notin P$. Since $\widehat{A^*} \times \widehat{A^*}$ is compact, the sequence (u_n, v_n) has a convergent subsequence. Let (u, v) be its limit. Since both P and P^c are closed and since the multiplication

in $\widehat{A^*}$ is continuous, one gets $usv \in P$ and $utv \notin P$. Therefore, $s \not\sim_P t$, which shows that $(\sim_P)^c$ is closed. Thus $\sim_P$ is clopen.

(2) implies (3). If $\sim_P$ is clopen, then for each $s \in \widehat{A^*}$, there exists an open neighborhood U of s such that $U \times U \subseteq \sim_P$. Therefore U is contained in the $\sim_P$-class of s. This proves that the $\sim_P$-classes form an open partition of $\widehat{A^*}$. By compactness, this partition is finite and thus P is recognisable. Moreover, since each $\sim_P$-class is open, the syntactic morphism of P is continuous.

(3) implies (1). Let $\pi : \widehat{A^*} \to M$ be the syntactic morphism of P. Since P is recognisable, M is finite. One has $P = \pi^{-1}(\pi(P))$ and since M is finite, $\pi(P)$ is clopen in M. Finally, since π is continuous, P is clopen in $\widehat{A^*}$. □

We now turn to languages of A^*.

Proposition 3.16. *If L is a language of A^*, then $L = \overline{L} \cap A^*$. Furthermore, the following conditions are equivalent:*

(1) *L is recognisable,*

(2) *$L = K \cap A^*$ for some clopen subset K of $\widehat{A^*}$,*

(3) *$\overline{L}$ is clopen in $\widehat{A^*}$,*

(4) *$\overline{L}$ is recognisable in $\widehat{A^*}$.*

Proof. The inclusion $L \subseteq \overline{L} \cap A^*$ is obvious. Let $u \in \overline{L} \cap A^*$ and let M be the syntactic monoid of $\{u\}$. Since M separates u from any word v different from u, one gets $r(u, v) \leqslant |M|$ if $u \neq v$. Let $(u_n)_{n\in\mathbb{N}}$ be a sequence of words of L converging to u. If $d(u_n, u) < 2^{-|M|}$, one has necessarily $u = u_n$ and thus $u \in L$.

(1) implies (2). If L is recognisable, there is a morphism φ from A^* onto a finite monoid M such that $L = \varphi^{-1}(\varphi(L))$. Let $K = \widehat{\varphi}^{-1}(\varphi(L))$. Since M is discrete, $\varphi(L)$ is a clopen subset of M and since $\widehat{\varphi}$ is continuous, K is also clopen. Moreover, φ and $\widehat{\varphi}$ coincide on A^* and thus $L = \widehat{\varphi}^{-1}(\varphi(L)) \cap A^* = K \cap A^*$.

(2) implies (3). Suppose that $L = K \cap A^*$ with K clopen. Since K is open and A^* is dense in $\widehat{A^*}$, $K \cap A^*$ is dense in K. Thus $\overline{L} = \overline{K \cap A^*} = K$. Thus $\overline{L}$ is clopen in $\widehat{A^*}$.

(3) implies (4) follows from Proposition 3.15.

(4) implies (1). Let $\widehat{\eta} : \widehat{A^*} \to F$ be the syntactic morphism of $\overline{L}$ and let $P = \widehat{\eta}(\overline{L})$. Let η be the restriction of $\widehat{\eta}$ to A^*. Then we have $L = \overline{L} \cap A^* = \widehat{\eta}^{-1}(P) \cap A^* = \eta^{-1}(P)$. Thus L is recognisable. □

We now describe the closure in $\widehat{A^*}$ of a recognisable language of A^*.

Proposition 3.17. *Let L be a recognisable language of A^* and let $u \in \widehat{A^*}$. The following conditions are equivalent:*

(1) *$u \in \overline{L}$,*

(2) *$\widehat{\varphi}(u) \in \varphi(L)$, for all morphisms φ from A^* onto a finite monoid,*

(3) *$\widehat{\varphi}(u) \in \varphi(L)$, for some morphism φ from A^* onto a finite monoid that recognises L,*

(4) *$\widehat{\eta}(u) \in \eta(L)$, where η is the syntactic morphism of L.*

Proof. (1) implies (2). Let φ be a morphism from A^* onto a finite monoid F and let $\widehat{\varphi}$ be its continuous extension to $\widehat{A^*}$. Then $\widehat{\varphi}(\overline{L}) \subseteq \overline{\widehat{\varphi}(L)}$ since $\widehat{\varphi}$ is

continuous, and $\overline{\widehat{\varphi}(L)} = \widehat{\varphi}(L) = \varphi(L)$ since F is discrete. Thus if $u \in \overline{L}$, then $\widehat{\varphi}(u) \in \varphi(L)$.

(2) implies (4) and (4) implies (3) are trivial.

(3) implies (1). Let φ be a morphism from A^* onto a finite monoid F recognising L. Let u_n be a sequence of words of A^* converging to u. Since $\widehat{\varphi}$ is continuous, $\widehat{\varphi}(u_n)$ converges to $\widehat{\varphi}(u)$. But since F is discrete, $\widehat{\varphi}(u_n)$ is actually ultimately equal to $\widehat{\varphi}(u)$. Thus for n large enough, one has $\widehat{\varphi}(u_n) = \widehat{\varphi}(u)$. It follows by (3) that $\varphi(u_n) = \widehat{\varphi}(u_n) \in \varphi(L)$ and since φ recognises L, we finally get $u_n \in \varphi^{-1}(\varphi(L)) = L$. Therefore $u \in \overline{L}$. □

Corollary 3.18. *Let L be a recognisable language of A^* and let η be its syntactic morphism. Then $\overline{L} = \widehat{\eta}^{-1}(\eta(L))$.*

Let $\mathrm{Clopen}(\widehat{A^*})$ denote the Boolean algebra of all clopen sets of $\widehat{A^*}$.

Theorem 3.19. *The maps $L \mapsto \overline{L}$ and $K \mapsto K \cap A^*$ define mutually inverse isomorphisms between the Boolean algebras $\mathrm{Rec}(A^*)$ and $\mathrm{Clopen}(\widehat{A^*})$. In particular, the following formulas hold, for all $L, L_1, L_2 \in \mathrm{Rec}(A^*)$:*

(1) $\overline{L^c} = (\overline{L})^c$,

(2) $\overline{L_1 \cup L_2} = \overline{L}_1 \cup \overline{L}_2$,

(3) $\overline{L_1 \cap L_2} = \overline{L}_1 \cap \overline{L}_2$.

Proof. Property (1) follows from Proposition 3.17. Indeed, let η be the syntactic morphism of L. Then since $L = \eta^{-1}(\eta(L))$ and $L^c = \eta^{-1}(\eta(L)^c)$, one has $\eta(L^c) = \eta(L)^c$. Therefore, one gets the following sequence of equalities:

$$\overline{L^c} = \widehat{\eta}^{-1}(\eta(L^c)) = \widehat{\eta}^{-1}(\eta(L)^c) = [\widehat{\eta}^{-1}(\eta(L))]^c = (\overline{L})^c$$

Property (2) is a general result of topology and (3) is a consequence of (1) and (2). □

Theorem 3.19 shows that the closure operator behaves nicely with respect to Boolean operations. It also behaves nicely with respect to uniformly continuous functions.

Theorem 3.20. *Let $\varphi : A^* \to B^*$ be a uniformly continuous function and L be a recognisable language of B^*. Then $\widehat{\varphi}^{-1}(\overline{L}) = \overline{\widehat{\varphi}^{-1}(L)} = \overline{\varphi^{-1}(L)}$.*

Proof. First $\varphi^{-1}(L) = \widehat{\varphi}^{-1}(L) \cap A^*$. It follows by Theorem 3.19 (3) that

$$\overline{\varphi^{-1}(L)} = \overline{\widehat{\varphi}^{-1}(L) \cap A^*} = \overline{\widehat{\varphi}^{-1}(L)} \cap \overline{A^*} = \overline{\widehat{\varphi}^{-1}(L)} \cap \widehat{A^*} = \overline{\widehat{\varphi}^{-1}(L)} \tag{3.2}$$

The inclusion

$$\overline{\widehat{\varphi}^{-1}(L)} \subseteq \widehat{\varphi}^{-1}(\overline{L}) \tag{3.3}$$

now follows from the continuity of $\widehat{\varphi}$. Applying (3.3) to the recognisable language L^c and using 3.19 (1), one gets

$$(\overline{\widehat{\varphi}^{-1}(L)})^c = \overline{(\widehat{\varphi}^{-1}(L))^c} = \overline{\widehat{\varphi}^{-1}(L^c)} \subseteq \widehat{\varphi}^{-1}(\overline{L^c}) = \widehat{\varphi}^{-1}(\overline{L}^c) = (\widehat{\varphi}^{-1}(\overline{L}))^c \tag{3.4}$$

The conjunction of (3.3) and (3.4) now gives the equality $\widehat{\varphi}^{-1}(\overline{L}) = \overline{\varphi^{-1}(L)}$. □

Theorem 3.20 applies in particular to morphisms between free monoids.

Corollary 3.21. *Every morphism of monoids $\varphi : A^* \to B^*$ is uniformly continuous. Furthermore, if L is a recognisable language of B^*, then $\widehat{\varphi}^{-1}(\overline{L}) = \overline{\widehat{\varphi}^{-1}(L)} = \overline{\varphi^{-1}(L)}$.*

Proof. If L is a recognisable language of B^*, then $\varphi^{-1}(L)$ is a recognisable language of A^*. It follows by Theorem 2.7 that φ is uniformly continuous. The result now follows from Theorem 3.20. □

Here is another useful application of Theorem 3.20.

Corollary 3.22. *Let L be a recognisable language of A^* and let $x, y \in A^*$. Then $\overline{x^{-1}Ly^{-1}} = x^{-1}\overline{L}y^{-1}$.*

Proof. Let $\varphi : A^* \to A^*$ be the map defined by $\varphi(u) = xuy$. The formula $\varphi^{-1}(L) = x^{-1}Ly^{-1}$ shows that if L is recognisable, then so is $\varphi^{-1}(L)$. It follows by Theorem 2.7 that φ is uniformly continuous. Furthermore, $\widehat{\varphi}^{-1}(\overline{L}) = \{u \in \widehat{A^*} \mid xuy \in \overline{L}\} = x^{-1}\overline{L}y^{-1}$. Applying Theorem 3.20, one gets $\widehat{\varphi}^{-1}(\overline{L}) = \overline{\varphi^{-1}(L)}$, that is, $x^{-1}\overline{L}y^{-1} = \overline{x^{-1}Ly^{-1}}$. □

4 Exercises

Exercise 1. A metric d on a space E is said to be an *ultrametric* if, for every $x, y, z \in E$, $d(x, z) \leqslant \max\{d(x, y), d(y, z)\}$. Show that if $d(x, y) \neq d(y, z)$, then $d(x, z) = \max\{d(x, y), d(y, z)\}$. Show that any open ball $B(x, \varepsilon)$ is clopen and that for any $y \in B(x, r)$, $B(x, r) = B(y, r)$.

5 Notes

Profinite algebra goes back at least to Birkhoff [16, p. 52-53], where he introduces topologies defined by congruences on abstract algebras. Profinite topologies for free groups were subsequently explored by M. Hall [56] and is now an important branch of group theory. The profinite approach has also been used to much profit in semigroup theory and in automata theory, in particular by Reutenauer [134] and by Almeida, who greatly developed the theory [2, 4, 5]. In particular, Almeida's book [4] is a major reference. See also the survey of Weil [175].

Chapter XI

Varieties

The definition of a variety of finite monoids is due to Eilenberg [42]. It is inspired by the definition of a Birkhoff variety of monoids (see Exercise 1) which applies to infinite monoids, while Eilenberg's definition is restricted to finite monoids. The word "variety" was coined after the term used in algebraic geometry to designate the solutions of a set of algebraic equations. This is no coincidence: a theorem of Birkhoff states that a Birkhoff variety can be described by a set of identities (see Exercise 2). The counterpart of Birkhoff's theorem for varieties of finite monoids was obtained by Reiterman [133] and independently by Banaschewski [10]. The statement is exactly the same: *any variety of finite monoids can be described by a set of identities*, but the definition of an identity is now different! For Birkhoff, an identity is a formal equality between words, but for Reiterman, it is a formal equality between profinite words. One must always topologize...

Warning. Since we are mostly interested in finite semigroups, the semigroups and monoids considered in this chapter are either finite or free, except in Exercises 1 and 2.

1 Varieties

A *variety of semigroups* is a class of semigroups $\mathbf{V}$ such that:

(1) if $S \in \mathbf{V}$ and if T is a subsemigroup of S, then $T \in \mathbf{V}$,

(2) if $S \in \mathbf{V}$ and if T is a quotient of S, then $T \in \mathbf{V}$,

(3) if $(S_i)_{i\in I}$ is a finite family of semigroups of $\mathbf{V}$, then $\prod_{i\in I} S_i$ is also in $\mathbf{V}$.

Conditions (1) and (2) can be replaced by a single condition: if $S \in \mathbf{V}$ and if T divides S, then $T \in \mathbf{V}$. Therefore, a variety of semigroups can be defined as a class of semigroups closed under division and finite products.

A more subtle point: condition (3) can be replaced by the conjunction of conditions (4) and (5):

(4) the trivial semigroup 1 belongs to $\mathbf{V}$,

(5) if S_1 and S_2 are semigroups of $\mathbf{V}$, then $S_1 \times S_2$ is also in $\mathbf{V}$.

Indeed, condition (4) is obtained by taking $I = \emptyset$ in (3).

Example 1.1.

(1) The class **S** of all semigroups forms a variety of semigroups.
(2) The smallest variety of semigroups is the trivial variety, consisting of the empty semigroup and of the semigroup with one element 1. This variety is denoted by **1**.

Varieties of monoids, varieties of ordered semigroups and varieties of ordered monoids are defined in the same way: it suffices to replace every occurrence of *semigroup* by *monoid* [*ordered semigroup*, *ordered monoid*] in the definition. We shall sometimes use the term "variety" as a shorthand for variety of semigroups [monoids, etc.]. Examples of varieties are deferred to Section 4.

Let $\mathcal{C}$ be a class of monoids [semigroups, ordered monoids, ordered semigroups]. The intersection of all varieties containing $\mathcal{C}$ is a variety, called the *variety generated* by $\mathcal{C}$ and denoted by $\langle\mathcal{C}\rangle$. The next proposition provides a more explicit description.

Proposition 1.1. *A monoid belongs to $\langle\mathcal{C}\rangle$ if and only if it divides a finite product of monoids of $\mathcal{C}$.*

Proof. Let **V** be the class of all monoids dividing a finite product of monoids of $\mathcal{C}$. Since a variety is closed under division and finite product, every monoid of **V** belongs to $\langle\mathcal{C}\rangle$. Therefore, it suffices to prove that **V** is a variety. Proposition II.3.14 shows that **V** is closed under product and Proposition II.3.9 shows that **V** is closed under division. Thus **V** is a variety. □

The *supremum* of two varieties $\mathbf{V}_1$ and $\mathbf{V}_2$ is the variety generated by $\mathbf{V}_1$ and $\mathbf{V}_2$. It is denoted by $\mathbf{V}_1 \vee \mathbf{V}_2$. A direct application of Proposition 1.1 gives the following characterisation.

Corollary 1.2. *A monoid belongs to $\mathbf{V}_1 \vee \mathbf{V}_2$ if and only if it divides a monoid of the form $M_1 \times M_2$, with $M_1 \in \mathbf{V}_1$ and $M_2 \in \mathbf{V}_2$.*

2 Free pro-V monoids

We introduced the free profinite monoid $\widehat{A^*}$ in Section X.2. A similar notion, the free pro-**V** monoid, can be defined for each variety of monoids [semigroups] **V**.

A monoid M *separates* two words u and v of the free monoid A^* if there exists a morphism φ from A^* onto M such that $\varphi(u) \neq \varphi(v)$. Let **V** be a variety of monoids. We set

$$r_{\mathbf{V}}(u,v) = \min\{\mathrm{Card}(M) \mid M \text{ is a monoid of } \mathbf{V} \text{ that separates } u \text{ and } v\}$$

and $d_{\mathbf{V}}(u,v) = 2^{-r_{\mathbf{V}}(u,v)}$, with the usual conventions $\min\emptyset = +\infty$ and $2^{-\infty} = 0$. We first establish some general properties of $d_{\mathbf{V}}$.

Proposition 2.3. *The following properties hold for every $u, v, w \in A^*$*

(1) $d_{\mathbf{V}}(u,v) = d_{\mathbf{V}}(v,u)$
(2) $d_{\mathbf{V}}(uw,vw) \leqslant d_{\mathbf{V}}(u,v)$ *and* $d_{\mathbf{V}}(wu,wv) \leqslant d_{\mathbf{V}}(u,v)$
(3) $d_{\mathbf{V}}(u,w) \leqslant \max\{d_{\mathbf{V}}(u,v), d_{\mathbf{V}}(v,w)\}$

Proof. The first assertion is trivial. A monoid of **V** separating uw and vw certainly separates u and v. Therefore $d_{\mathbf{V}}(uw, vw) \leqslant d_{\mathbf{V}}(u, v)$, and dually, $d_{\mathbf{V}}(wu, wv) \leqslant d_{\mathbf{V}}(u, v)$.

Let M be a monoid of **V** separating u and w. Then M separates either u and v, or v and w. It follows that $\min(r_{\mathbf{V}}(u, v), r_{\mathbf{V}}(v, w)) \leqslant r_{\mathbf{V}}(u, w)$ and hence $d_{\mathbf{V}}(u, w) \leqslant \max\{d_{\mathbf{V}}(u, v), d_{\mathbf{V}}(v, w)\}$. □

In the general case, $d_{\mathbf{V}}$ is not always a metric, because one may have $d_{\mathbf{V}}(u, v) = 0$ even if $u \neq v$. For instance, if **V** is the variety of commutative monoids, $d_{\mathbf{V}}(ab, ba) = 0$, since there is no way to separate ab and ba by a commutative monoid. To work around this difficulty, we first observe that, by Proposition 2.3, the relation $\sim_{\mathbf{V}}$ defined by

$$u \sim_{\mathbf{V}} v \text{ if and only if } d_{\mathbf{V}}(u, v) = 0$$

is a congruence on A^*. Then Proposition X.2.9 can be generalised as follows.

Proposition 2.4.

(1) *The function $d_{\mathbf{V}}$ is an ultrametric on $A^*/\sim_{\mathbf{V}}$.*

(2) *The product on $A^*/\sim_{\mathbf{V}}$ is uniformly continuous for this metric.*

Proof. (1) follows directly from Proposition 2.3, since $d_{\mathbf{V}}(u, v) = 0$ implies $u \sim_{\mathbf{V}} v$ by definition. We use the same proposition to obtain the relation

$$d_{\mathbf{V}}(uv, u'v') \leqslant \max\{d_{\mathbf{V}}(uv, uv'), d_{\mathbf{V}}(uv', u'v')\} \leqslant \max\{d_{\mathbf{V}}(v, v'), d_{\mathbf{V}}(u, u')\}$$

which proves (2). □

The completion of the metric space $(A^*/\sim_{\mathbf{V}}, d_{\mathbf{V}})$ is denoted by $\widehat{F}_{\mathbf{V}}(A)$ and called the *free pro-***V** *monoid* on A. Its main properties are presented in the next three propositions. Let $\pi_{\mathbf{V}}$ be the natural morphism from A^* onto $A^*/\sim_{\mathbf{V}}$.

Proposition 2.5. *The following properties hold for each finite alphabet A:*

(1) *The monoid $\widehat{F}_{\mathbf{V}}(A)$ is compact.*

(2) *There is a unique surjective uniformly continuous morphism from $\widehat{A^*}$ onto $\widehat{F}_{\mathbf{V}}(A)$ extending $\pi_{\mathbf{V}}$.*

Proof. (1) Since $\widehat{F}_{\mathbf{V}}(A)$ is complete, it suffices to verify that, for every $n > 0$, A^* is covered by a finite number of open balls of radius $< 2^{-n}$. Consider the congruence $\sim_n$ defined on A^* by

$$u \sim_n v \text{ if and only if } \varphi(u) = \varphi(v) \text{ for every morphism } \varphi \text{ from } A^* \text{ onto a monoid of size } \leqslant n \text{ of } \mathbf{V}.$$

Since A is finite, there are only finitely many morphisms from A^* onto a monoid of size $\leqslant n$, and thus $\sim_n$ is a congruence of finite index. Furthermore, $d_{\mathbf{V}}(u, v) < 2^{-n}$ if and only if u and v cannot be separated by a monoid of **V** of size $\leqslant n$, i.e. are $\sim_n$-equivalent. It follows that the $\sim_n$-classes are open balls of radius $< 2^{-n}$ and cover A^*.

(2) Since $d_{\mathbf{V}}(u, v) \leqslant d(u, v)$, $\pi_{\mathbf{V}}$ is uniformly continuous, and since $\widehat{A^*}$ is compact, Proposition X.1.1 shows that it can be extended (in a unique way) to a uniformly continuous morphism from $\widehat{A^*}$ onto $\widehat{F}_{\mathbf{V}}(A)$. □

The monoid $\widehat{F}_{\mathbf{V}}(A)$ has the following universal property:

Proposition 2.6. *For each mapping φ from A to a monoid M of $\mathbf{V}$, there is a unique uniformly continuous morphism $\widehat{\varphi} : \widehat{F}_{\mathbf{V}}(A) \to M$ such that, for all $a \in A$, $\varphi(a) = \widehat{\varphi}(\pi_{\mathbf{V}}(a))$.*

Proof. Let φ be a continuous morphism from $\widehat{A^*}$ to a monoid M of $\mathbf{V}$. Up to replacing M by $\varphi(\widehat{A^*})$, we may assume that φ is surjective. Since A^* is dense in $\widehat{A^*}$, and M is discrete, the restriction of φ to A^* is also surjective. Furthermore, since $u \sim_{\mathbf{V}} v$ implies $\varphi(u) = \varphi(v)$, Proposition II.3.22 shows that there is a surjective morphism π from $A^*/\sim_{\mathbf{V}}$ onto M such that $\varphi = \pi \circ \pi_{\mathbf{V}}$. We claim that this morphism is uniformly continuous. Indeed if $d_{\mathbf{V}}(u, v) < 2^{-|M|}$, then u and v cannot be separated by M, and hence $\varphi(u) = \varphi(v)$. Since $A^*/\sim_{\mathbf{V}}$ is dense in $\widehat{F}_{\mathbf{V}}(A)$, π can be extended by continuity to a surjective morphism from $\widehat{F}_{\mathbf{V}}(A)$ onto M. Thus M is a quotient of $\widehat{F}_{\mathbf{V}}(A)$. □

Proposition 2.7. *A finite A-generated monoid belongs to $\mathbf{V}$ if and only if it is a continuous quotient of $\widehat{F}_{\mathbf{V}}(A)$.*

Proof. If M is an A-generated monoid of $\mathbf{V}$, there exists a surjective morphism φ from A^* onto M. Following the argument used in the proof of Proposition 2.6, if $d_{\mathbf{V}}(u, v) < 2^{-|M|}$, then $\varphi(u) = \varphi(v)$, and thus φ is uniformly continuous with respect to $d_{\mathbf{V}}$. By Proposition X.1.1, φ can be extended to a uniformly continuous morphism from $\widehat{F}_{\mathbf{V}}(A)$ onto M.

Conversely, assume that M is a finite quotient of $\widehat{F}_{\mathbf{V}}(A)$ and let $\pi : \widehat{F}_{\mathbf{V}}(A) \to M$ be a surjective morphism. The set

$$D = \{(u, v) \in \widehat{F}_{\mathbf{V}}(A) \times \widehat{F}_{\mathbf{V}}(A) \mid \pi(u) = \pi(v)\}$$

is the inverse image under π of the diagonal of $M \times M$, and since M is discrete and π is continuous, it is a clopen subset of $\widehat{F}_{\mathbf{V}}(A) \times \widehat{F}_{\mathbf{V}}(A)$. Let $\mathcal{M}$ be the class of all morphisms from $\widehat{F}_{\mathbf{V}}(A)$ onto a monoid of $\mathbf{V}$. For each $\varphi \in \mathcal{M}$, let

$$C_\varphi = \{(u, v) \in \widehat{F}_{\mathbf{V}}(A) \times \widehat{F}_{\mathbf{V}}(A) \mid \varphi(u) \neq \varphi(v)\}$$

Each C_φ is open by continuity of φ. Furthermore, if (u, v) does not belong to any C_φ, then u and v cannot be separated by any monoid of $\mathbf{V}$ and hence $d_{\mathbf{V}}(u, v) = 0$, which gives $u = v$ and $\pi(u) = \pi(v)$. It follows that the family $D \cup (C_\varphi)_{\varphi \in \mathcal{M}}$ is a covering of $\widehat{F}_{\mathbf{V}}(A) \times \widehat{F}_{\mathbf{V}}(A)$ by open sets, and since $\widehat{F}_{\mathbf{V}}(A)$ is compact, it admits a finite subcovering, say $D \cup (C_\varphi)_{\varphi \in \mathcal{F}}$. Therefore, if $\varphi(u) = \varphi(v)$ for each $\varphi \in \mathcal{F}$, then $\pi(u) = \pi(v)$. Consequently M is a quotient of a submonoid of the finite monoid $\prod_{\varphi \in \mathcal{F}} \varphi(\widehat{F}_{\mathbf{V}}(A))$ and thus belongs to $\mathbf{V}$. □

Given a variety $\mathbf{V}$, it is in general a hard problem to describe the structure of $\widehat{F}_{\mathbf{V}}(A)$. However, if $\mathbf{V}$ is generated by a single monoid M, then $\widehat{F}_{\mathbf{V}}(A)$ has a simple structure. Let M^A be the set of all functions from A to M. Each function $\varphi : A \to M$ extends in a unique way to a monoid morphism from A^* to M, also denoted by φ. Denote by M^{M^A} the monoid of all functions from M^A to M under pointwise multiplication.

For each letter $a \in A$, let $\tilde{a}$ be the function from M^A to M defined by $\tilde{a}(\varphi) = \varphi(a)$. This defines a function $a \mapsto \tilde{a}$ from A to M^{M^A}, which can be

extended to a morphism $u \mapsto \tilde{u}$ from A^* to M^{M^A}. We claim that, for any function $\varphi : A \to M$, one has $\tilde{u}(\varphi) = \varphi(u)$ for each word $u \in A^*$. Indeed, if $u = a_1 \cdots a_n$, one gets by definition

$$\tilde{u}(\varphi) = \tilde{a}_1(\varphi) \cdots \tilde{a}_n(\varphi) = \varphi(a_1) \cdots \varphi(a_n) = \varphi(u)$$

The image of A^* under the map $u \mapsto \tilde{u}$ is a submonoid F of M^{M^A}.

Proposition 2.8. *If* $\mathbf{V}$ *is generated by a single monoid* M*, then* $\widehat{F}_{\mathbf{V}}(A)$ *is equal to* F*. In particular,* $\widehat{F}_{\mathbf{V}}(A)$ *is a submonoid of* M^{M^A} *and hence is finite.*

Proof. It suffices to prove that $u \sim_{\mathbf{V}} v$ is equivalent to $\tilde{u} = \tilde{v}$. Recall that $u \sim_{\mathbf{V}} v$ if and only if, for each morphism $\varphi : A^* \to N$, where $N \in \mathbf{V}$, one has $\varphi(u) = \varphi(v)$.

Suppose that $u \sim_{\mathbf{V}} v$ and let $\varphi : A \to M$ be a function, which extends to a morphism from A^* to M. Since $M \in \mathbf{V}$, one has $\varphi(u) = \varphi(v)$ and thus $\tilde{u}(\varphi) = \tilde{v}(\varphi)$. It follows that $\tilde{u} = \tilde{v}$.

Suppose now that $\tilde{u} = \tilde{v}$ and let φ be a morphism from A^* onto a monoid N of $\mathbf{V}$. Since $\mathbf{V}$ is generated by M, N divides a power of M. Therefore there exist a positive integer n, a submonoid T of M^n and a surjective morphism $\pi : T \to N$. By Corollary II.5.30, there exists a morphism $\alpha : A^* \to T$ such that $\varphi = \pi \circ \alpha$. Denoting by π_i the i-th projection from M^n onto M and setting $\alpha_i = \pi_i \circ \alpha$, we get $\alpha(u) = (\alpha_1(u), \ldots, \alpha_n(u))$. Now since α_i is a morphism from A^* to M, one has $\alpha_i(u) = \tilde{u}(\alpha_i)$. Since $\tilde{u} = \tilde{v}$, it follows that $\alpha(u) = \alpha(v)$ and finally $\varphi(u) = \varphi(v)$. Therefore $u \sim_{\mathbf{V}} v$, which concludes the proof. □

A variety such that $\widehat{F}_{\mathbf{V}}(A)$ is finite for each alphabet A is called *locally finite*. It is tempting to guess, in view of Proposition 2.8, that every locally finite variety is generated by a single monoid. However, this is not the case: one can show that the variety of idempotent monoids is locally finite but not finitely generated.

3 Identities

3.1 What is an identity?

Let A be a finite alphabet and let u and v be two profinite words of $\widehat{A}^*$. A monoid M *satisfies the profinite identity* $u = v$ if, for each monoid morphism $\varphi : A^* \to M$, one has $\widehat{\varphi}(u) = \widehat{\varphi}(v)$. Similarly, an ordered monoid $(M, \leqslant)$ satisfies the *profinite identity* $u \leqslant v$ if, for each monoid morphism $\varphi : A^* \to M$, one has $\widehat{\varphi}(u) \leqslant \widehat{\varphi}(v)$. An identity $u = v$ $[u \leqslant v]$ where u and v are words is sometimes called an *explicit identity*.

One can give similar definitions for [ordered] semigroups, by taking two nonempty profinite words u and v. A semigroup S *satisfies the identity* $u = v$ if, for each semigroup morphism $\varphi : A^+ \to S$, one has $\widehat{\varphi}(u) = \widehat{\varphi}(v)$ $[\widehat{\varphi}(u) \leqslant \widehat{\varphi}(v)]$. The context is usually sufficient to know which kind of identities is considered. In case of ambiguity, one can utilise the more precise terms [*ordered*] *monoid identity* or [*ordered*] *semigroup identity*.

Formally, a profinite identity is a pair of profinite words. But in practice, one often thinks directly in terms of elements of a monoid. Consider for instance

the explicit identity $xy^3z = yxyzy$. A monoid M satisfies this identity if, for each morphism $\gamma : \{x, y, z\}^* \to M$, one has $\gamma(xy^3z) = \gamma(yxyzy)$. Setting $\gamma(x) = s$, $\gamma(y) = t$ and $\gamma(z) = u$, this can be written as $st^3u = tstut$. Since this equality should hold for **any** morphism γ, it is equivalent to require that, for **all** $s, t, u \in M$, one has $st^3u = tstut$. Now, the change of variables is unnecessary, and one can write directly that M satisfies the identity $xy^3z = yxyzy$ if and only if, for all $x, y, z \in M$, one has $xy^3z = yxyzy$. Similarly, a monoid is commutative if it satisfies the identity $xy = yx$.

It is also possible to directly interpret any ω-term in a finite semigroup [monoid] S. As it was already mentioned, the key idea is to think of ω as the exponent of S. For instance, an identity like $x^\omega y^\omega = y^\omega x^\omega$ can be readily interpreted by saying that idempotents commute in S. The identity $x^\omega y x^\omega = x^\omega$ means that, for every $e \in E(S)$ and for every $s \in S$, $ese = e$. Propositions 4.23 and 4.27 provide other enlighting examples.

3.2 Properties of identities

Let us now state some elementary but important properties of identities. We consider identities of ordered monoids, but the result can be readily adapted to the other types of identities.

Proposition 3.9. *Let u and v be profinite words on the alphabet A and let M be an ordered monoid satisfying the identity $u \leqslant v$. Then for all $x, y \in \widehat{A^*}$, M satisfies the identity $xuy \leqslant xvy$. Furthermore, for each morphism $\gamma : A^* \to B^*$, M satisfies the identity $\widehat{\gamma}(u) \leqslant \widehat{\gamma}(v)$.*

Proof. Let $\varphi : A^* \to M$ be a monoid morphism. Since M satisfies the identity $u \leqslant v$, one has $\widehat{\varphi}(u) \leqslant \widehat{\varphi}(v)$. Since $\widehat{\varphi}$ is a morphism, it follows that $\widehat{\varphi}(xuy) \leqslant \widehat{\varphi}(xvy)$ for all $x, y \in \widehat{A^*}$. Therefore, M satisfies the identity $xuy \leqslant xvy$.

Let $\gamma : A^* \to B^*$ and $\alpha : B^* \to M$ be morphisms. Then $\alpha \circ \gamma$ is a morphism from A^* to M. Since M satisfies the identity $u \leqslant v$ and since $\widehat{\alpha \circ \gamma} = \widehat{\alpha} \circ \widehat{\gamma}$, one has $\widehat{\alpha}(\widehat{\gamma}(u)) \leqslant \widehat{\alpha}(\widehat{\gamma}(v))$. Therefore M satisfies the identity $\widehat{\gamma}(u) \leqslant \widehat{\gamma}(v)$. □

It is a common practice to implicitly use the second part of Proposition 3.9 without giving the morphism γ explicitly, but simply by substituting a profinite word for a letter. For instance, one can prove that the monoid identity

$$xyx = x \tag{3.1}$$

implies the identity $x^2 = x$ by taking $y = 1$ in (3.1). Similarly, the identity $x^\omega y x^\omega = x^\omega$ implies the identity $x^{\omega+1} = x^\omega$ by taking $y = x$, since $x^\omega x x^\omega = x^\omega x^\omega x = x^\omega x$.

3.3 Reiterman's theorem

Given a set E of profinite identities, the class of monoids [semigroups, ordered monoids, ordered semigroups] defined by E is the class of all monoids [semigroups, ordered monoids, ordered semigroups] satisfying all the identities of E and is denoted by $[\![E]\!]$.

The main result of this section, Reiterman's theorem (Theorem 3.13), states that varieties can be characterised by profinite identities. We first establish the easy part of this result.

Proposition 3.10. *Let E be a set of identities. Then $[\![E]\!]$ is a variety of monoids [semigroups, ordered monoids, ordered semigroups].*

Proof. We treat only the case of ordered monoids, but the other cases are similar. Since varieties are closed under intersection, it suffices to prove the result when E consists of a single identity, say $u \leqslant v$. Let M be an ordered monoid satisfying this identity. Clearly, every submonoid of M satisfies the same identity.

Let N be a quotient of M and let $\pi : M \to N$ be a surjective morphism. We claim that N also satisfies $u \leqslant v$. Indeed, if $\varphi : A^* \to N$ is a morphism, there exists by Corollary II.5.30 a morphism $\psi : A^* \to M$ such that $\varphi = \pi \circ \psi$. Since M satisfies the identity $u \leqslant v$, one gets $\widehat{\psi}(u) \leqslant \widehat{\psi}(v)$ and thus $\pi(\widehat{\psi}(u)) \leqslant \pi(\widehat{\psi}(v))$. Finally, since $\widehat{\varphi} = \pi \circ \widehat{\psi}$, one obtains $\widehat{\varphi}(u) \leqslant \widehat{\varphi}(v)$, which proves the claim.

Finally, let $(M_i)_{i\in I}$ be a finite family of ordered monoids satisfying the identity $u \leqslant v$. We claim that their product $M = \prod_{i\in I} M_i$ also satisfies this identity. Indeed, let π_i denote the projection from M onto M_i and let φ be a morphism from A^* to M. Since $\pi_i \circ \varphi$ is a morphism from A^* to M_i and since $\widehat{\pi_i \circ \varphi} = \pi_i \circ \widehat{\varphi}$, one has $\pi_i \circ \widehat{\varphi}(u) \leqslant \pi_i \circ \widehat{\varphi}(v)$. As this holds for each i, one has $\widehat{\varphi}(u) \leqslant \widehat{\varphi}(v)$. This proves the claim and concludes the proof. □

A variety $\mathbf{V}$ satisfies a given identity if every monoid of $\mathbf{V}$ satisfies this identity. We also say in this case that the given identity is an identity of $\mathbf{V}$. Identities of $\mathbf{V}$ are closely related to free pro-$\mathbf{V}$ monoids.

Proposition 3.11. *Let A be a finite alphabet. Given two profinite words u and v of $\widehat{A^*}$, $u = v$ is an identity of $\mathbf{V}$ if and only if $\pi_{\mathbf{V}}(u) = \pi_{\mathbf{V}}(v)$.*

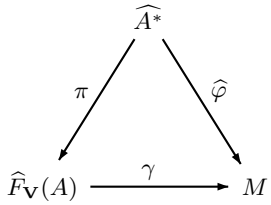
Proof. If $u = v$ is an identity of $\mathbf{V}$, then u and v cannot be separated by any monoid of $\mathbf{V}$. Thus $d_{\mathbf{V}}(u, v) = 0$, $u \sim_{\mathbf{V}} v$ and $\pi_{\mathbf{V}}(u) = \pi_{\mathbf{V}}(v)$. Conversely if $\pi_{\mathbf{V}}(u) = \pi_{\mathbf{V}}(v)$, then by Proposition 2.5, $\varphi(u) = \varphi(v)$ for every continuous morphism φ from $\widehat{A^*}$ to a monoid of $\mathbf{V}$, and thus $u = v$ is an identity of $\mathbf{V}$. □

Corollary 3.12. *Let $\mathbf{V}$ and $\mathbf{W}$ be two varieties of monoids satisfying the same identities on the alphabet A. Then $\widehat{F}_{\mathbf{V}}(A)$ and $\widehat{F}_{\mathbf{W}}(A)$ are isomorphic.*

In particular, an identity of a monoid of $\mathbf{V}$ can be given as a pair (u, v) of elements of $\widehat{F}_{\mathbf{V}}(A)$. Given a set E of identities, let $[\![E]\!]$ denote the class of monoids satisfying all the identities of E. Reiterman's theorem states that these classes are exactly the varieties of monoids [semigroups, ordered monoids, ordered semigroups].

Theorem 3.13 (Reiterman)**.** *A class of [ordered] monoids [semigroups] is a variety if and only if it can be defined by a set of profinite identities.*

Proof. The first part of the theorem follows from Proposition 3.10. Let now $\mathbf{V}$ be a variety of monoids. Let E be the class of all identities which are satisfied by every monoid of $\mathbf{V}$ and let $\mathbf{W} = [\![E]\!]$. Clearly $\mathbf{V} \subseteq \mathbf{W}$. Let $M \in \mathbf{W}$. Since M is finite, there exists a finite alphabet A and a surjective morphism $\varphi : A^* \to M$ which can be extended to a uniformly continuous morphism from $\widehat{A^*}$ onto M. Let $\pi : \widehat{A^*} \to \widehat{F}_{\mathbf{V}}(A)$ be the natural morphism and let $u, v \in \widehat{A^*}$. By Proposition 3.11, if $\pi(u) = \pi(v)$, then $u = v$ is an identity of $\mathbf{V}$ and thus, is satisfied by M. In particular, $\pi(u) = \pi(v)$ implies $\varphi(u) = \varphi(v)$ and by Proposition II.3.22, there is a morphism $\gamma : \widehat{F}_{\mathbf{V}}(A) \to M$ such that $\gamma \circ \pi = \widehat{\varphi}$.



We claim that γ is continuous. Since M is discrete, it suffices to prove that if $m \in M$, then $\gamma^{-1}(m)$ is closed. But $\gamma^{-1}(m) = \pi(\widehat{\varphi}^{-1}(m))$ and since $\widehat{\varphi}$ is continuous, $\widehat{\varphi}^{-1}(m)$ is closed and hence compact. It follows that $\pi(\widehat{\varphi}^{-1}(m))$ is compact and hence closed which proves the claim. Therefore M is a continuous quotient of $\widehat{F}_{\mathbf{V}}(A)$ and by Proposition 2.7, M is in $\mathbf{V}$. Thus $\mathbf{V} = \mathbf{W}$. □

4 Examples of varieties

We now illustrate Reiterman's theorem by giving identities defining various varieties.

4.1 Varieties of semigroups

Nilpotent semigroups

A semigroup S is *nilpotent* if it has a zero and $S^n = 0$ for some positive integer n. Recall that S^n denotes the set of all products of the form $s_1 \cdots s_n$, with $s_1, \ldots, s_n \in S$. Equivalent definitions are given below.

Proposition 4.14. *Let S be a nonempty semigroup. The following conditions are equivalent:*

(1) *S is nilpotent,*

(2) *S satisfies the identity $x_1 \cdots x_n = y_1 \cdots y_n$, where $n = |S|$,*

(3) *for every $e \in E(S)$ and every $s \in S$, one has $es = e = se$,*

(4) *S has a zero which is the only idempotent of S.*

Proof. (1) implies (2). This follows immediately from the definition.

(2) implies (3). Let $s \in S$ and $e \in E(S)$. Taking $x_1 = s$ and $x_2 = \ldots = x_n = y_1 = \ldots = y_n = e$, one gets $s = e$ if $n = 1$ and $se = e$ if $n > 1$. Therefore $se = e$ in all cases. Similarly $es = e$, which proves (3).

(3) implies (4). Since S is a nonempty semigroup, it contains an idempotent by Corollary II.6.32. Moreover, by (3), every idempotent of S is a zero of S, but Proposition II.1.2 shows that a semigroup has at most one zero.

(4) implies (1). Denote by 0 the zero of S. Then by Corollary II.6.35, $S^n = SE(S)S = S\{0\}S = 0$. □

Nilpotent semigroups form a variety, denoted by $\mathbf{N}$, and defined by the identities $x^\omega y = x^\omega = yx^\omega$.

Lefty and righty trivial semigroups

A semigroup is called *lefty* [*righty*] *trivial* if, for every idempotent e of S, the semigroup eS [Se] is trivial. Equivalently, e is a left [right] zero of S, that is, for every $s \in S$, one has $es = e$ [$se = e$]. Equivalent definitions are possible:

Proposition 4.15. *Let S be a nonempty semigroup and let I be its minimal ideal. Let $n = |S|$. The following conditions are equivalent:*

(1) *S is lefty trivial,*

(2) *I is a left zero semigroup and is equal to $E(S)$,*

(3) *I is a left zero semigroup and $S^n = I$,*

(4) *S satisfies the identity $x_1 \cdots x_n x = x_1 \cdots x_n$.*

Proof. (1) implies (2). Let $s \in I$. Then for every $e \in E(S)$, one has $e = es \in I$ since I is an ideal and therefore $E(S) \subseteq I$. Moreover if $e \in E(S)$ and $s, t \in S^1$, $(set)(set) = se(tset) = se = set$ since e is a left zero of S. It follows that $E(S)$ is a nonempty ideal of S contained in I and hence $E(S) = I$ since I is minimal. Furthermore, since every idempotent of S is a left zero, I is a left zero semigroup.

(2) implies (3). By Corollary II.6.35, $S^n = SE(S)S$. If $I = E(S)$, this gives $S^n = SIS = I$.

(3) implies (4). Suppose that (3) holds and let $x, x_1, \ldots, x_n \in S$. Then $x_1 \cdots x_n \in I$ and since I is an ideal, $x_1 \cdots x_n x \in I$. Now since I is a left zero semigroup,

$$x_1 \cdots x_n = (x_1 \cdots x_n)(x_1 \cdots x_n x) = (x_1 \cdots x_n)(x_1 \cdots x_n)x = (x_1 \cdots x_n)x$$

which proves (4).

(4) implies (1). Let $s \in S$ and $e \in E(S)$. Taking $x = s$ and $x_1 = \ldots = x_n = e$, one gets $es = e$ and hence S is lefty trivial. □

Note that, in a lefty trivial semigroup, all regular elements are in the minimal ideal. Furthermore, all nonregular $\mathcal{D}$-classes are trivial. Let us state the dual result for the righty trivial semigroups.

Proposition 4.16. *Let S be a nonempty semigroup and let I be its minimal ideal. Let $n = |S|$. The following conditions are equivalent:*

(1) *S is righty trivial,*

(2) *I is a right zero semigroup and is equal to $E(S)$,*

(3) *I is a right zero semigroup and $S^n = I$,*

(4) *S satisfies the identity $xx_1 \cdots x_n = x_1 \cdots x_n$.*

The lefty [righty] trivial semigroups form a variety of finite semigroups, denoted by $\ell\mathbf{1}$ [r$\mathbf{1}$], and defined by the identity $x^\omega y = x^\omega$ [$yx^\omega = x^\omega$]. For historical reasons, a different notation is often used in the literature. The variety $\ell\mathbf{1}$ is denoted by $\mathbf{K}$ and the variety r$\mathbf{1}$ by $\mathbf{D}$.

Locally trivial semigroups

A semigroup is called *locally trivial* if for every $s \in S$ and $e \in E(S)$, $ese = e$. Equivalent definitions are given below.

Proposition 4.17. *Let S be a nonempty semigroup and let $n = |S|$. The following conditions are equivalent:*

(1) *S is locally trivial,*

(2) *the minimal ideal of S is equal to $E(S)$,*

(3) *for every $e, f \in E(S)$ and every $s \in S$, one has $esf = ef$,*

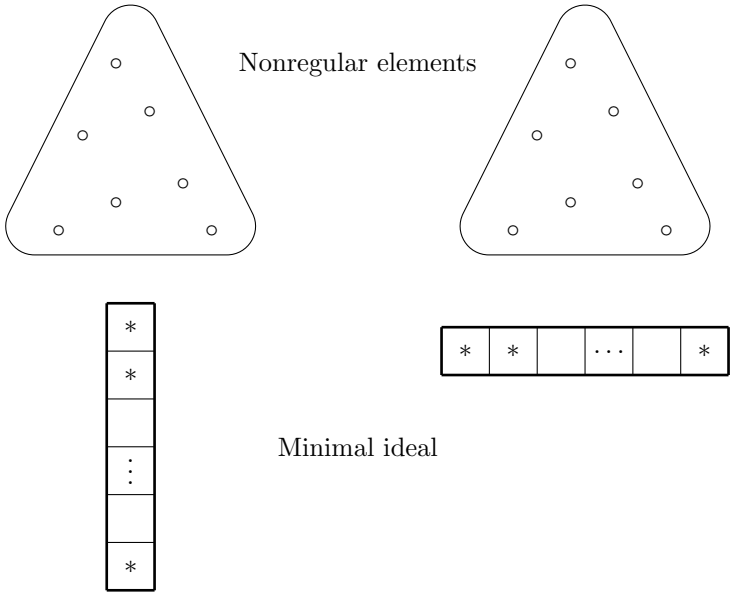


Figure 4.1. A lefty trivial semigroup (on the left) and a righty trivial semigroup (on the right).

(4) *S satisfies the identity $x_1 \cdots x_n x x_1 \cdots x_n = x_1 \cdots x_n$.*

Proof. (1) implies (2). Let $s \in I$. Then for every $e \in E(S)$, one has $e = ese \in I$ since I is an ideal and therefore $E(S) \subseteq I$. Moreover if $e \in E(S)$ and $s, t \in S^1$, $(set)(set) = se(ts)et = set$ and thus set is idempotent. It follows that $E(S)$ is a nonempty ideal of S contained in I and hence $E(S) = I$.

(2) implies (3). Suppose that $I = E(S)$. Then I is an aperiodic semigroup, which is also simple by Proposition V.4.37. Let $s \in S$ and let $e, f \in E(S)$. Then $e, f, esf \in I$ since I is an ideal. It follows by Corollary V.3.34 that $esf = eesff = e(esf)f = ef$.

(3) implies (4). By Corollary II.6.35, $S^n = SE(S)S$. Let $x, x_1, \ldots, x_n \in S$. Then $x_1 \cdots x_n \in S^n$ and hence $x_1 \cdots x_n = set$ for some $s, t \in S$ and $e \in E(S)$. It follows by (3)

$$x_1 \cdots x_n x x_1 \cdots x_n = (set)x(set) = s(etxse)t = set = x_1 \cdots x_n$$

which gives (4).

(4) implies (1). Let $s \in S$ and $e \in E(S)$. Taking $x = s$ and $x_1 = \ldots = x_n = e$, one gets $ese = e$ and hence S is locally trivial. □

Thus the locally trivial semigroups form a variety of finite semigroups, denoted by $\mathbb{L}\mathbf{1}$, and defined by the identity $x^\omega y x^\omega = x^\omega$.

Locally groups

A semigroup is *locally a group* if for every $e \in E(S)$, the semigroup eSe is a group. Equivalent definitions are given below.

Proposition 4.18. *Let S be a nonempty semigroup. The following conditions are equivalent:*

(1) *S is locally a group,*

(2) *every idempotent of S belongs to the minimal ideal of S,*

(3) *S satisfies the identity $(x^\omega y x^\omega)^\omega = x^\omega$,*

Proof. Let I be the minimal ideal of S and let n be the exponent of S.

(1) implies (2). Let $s \in I$ and let $e \in E(S)$. Since eSe is a group whose identity is e, one has $(ese)^n = e$. Now, I is an ideal and thus $e \in I$. It follows that $E(S)$ is a subset of I.

(2) implies (3). Suppose that $E(S) \subseteq I$. By Proposition V.4.37, I is a simple semigroup. Let $s \in S$ and let $e \in E(S)$. Since I is an ideal, one has $e, ese \in I$ and thus $e \mathcal{J}\ ese$. Since $ese \leqslant_{\mathcal{R}} e$ and $ese \leqslant_{\mathcal{L}} e$, Theorem V.1.9 shows that $ese\ \mathcal{H}\ e$. Therefore $(ese)^n = e$ and S satisfies the identity $(x^\omega y x^\omega)^\omega = x^\omega$.

(3) implies (1). Let $s \in S$ and $e \in E(S)$. Taking $x = e$ and $y = s$, one gets $(ese)^\omega = e$, which shows that S is locally a group. □

Semigroups which are locally a group form a variety of finite semigroups, denoted by $\mathbb{L}\mathbf{G}$ and defined by the identity $(x^\omega y x^\omega)^\omega = x^\omega$.

Simple semigroups

Proposition 4.19. *A semigroup is simple if and only if it satisfies the identities $x^{\omega+1} = x$ and $(xyx)^\omega = x^\omega$.*

Proof. Let S be a simple semigroup. By Proposition V.3.32, S has a single $\mathcal{D}$-class, which is a union of groups. Let $x, y \in S$. One has $x^{\omega+1} \leqslant_{\mathcal{H}} x$ and $(xyx)^\omega \leqslant_{\mathcal{H}} x$ and hence by Theorem V.1.9 (3) and (4), $x^{\omega+1}\ \mathcal{H}\ x$ and $(xyx)^\omega\ \mathcal{H}\ x^\omega$. It follows immediately that $(xyx)^\omega = x^\omega$ since an $\mathcal{H}$-class contains a unique idempotent and $x^{\omega+1} = x$ since x^ω is the identity of the $\mathcal{H}$-class containing x and $x^{\omega+1}$.

Conversely, suppose that S satisfies the two identities

$$x^{\omega+1} = x \text{ and } (xyx)^\omega = x^\omega$$

The first identity shows that every element x is $\mathcal{H}$-equivalent to x^ω and to $x^{\omega+1}$ and hence belongs to the maximal group whose identity is x^ω. In particular, all the elements of S are regular. Furthermore, if $x, y \in S$, one has $x\ \mathcal{J}\ x^\omega = (xyx)^\omega \leqslant_{\mathcal{J}} xyx \leqslant_{\mathcal{J}} y$. It follows that $x \leqslant_{\mathcal{J}} y$ and by duality $y \leqslant_{\mathcal{J}} x$. It follows that all elements of S are $\mathcal{J}$-equivalent. □

Simple semigroups form a variety of semigroups, usually denoted by **CS**.

4.2 Varieties of monoids

Groups

The next proposition shows that groups form a variety of monoids, denoted by **G**.

Proposition 4.20. *The class of all groups is a variety of monoids, defined by the identity $x^\omega = 1$.*

Proof. Finite groups are closed under quotients and finite direct products and it follows from Proposition II.3.13 that a submonoid of a group is a group.

Since a monoid is a group if and only if its unique idempotent is 1, the identity $x^\omega = 1$ characterises the variety of finite groups. □

Subvarieties of **G** include the variety **Gcom** of commutative groups and, for each prime number p, the variety $\mathbf{G}_p$ of p-groups. Recall that a group is a p-group if its order is a power of p.

Commutative monoids

A monoid is commutative if and only if it satisfies the identity $xy = yx$. Therefore, the commutative monoids form a variety of monoids, denoted by **Com**.

Proposition 4.21. *Every commutative monoid is a quotient of the product of its monogenic submonoids.*

Proof. Let M be a commutative monoid and let N be the product of its monogenic submonoids. Let $\varphi : N \to M$ be the morphism which maps each element of N to the product of its coordinates. Then φ is clearly surjective and thus M is a quotient of N. □

Aperiodic monoids

Observe that a monoid is *aperiodic* if and only if it satisfies the identity $x^\omega = x^\omega x$, which can also be written, by abuse of notation, $x^\omega = x^{\omega+1}$. Thus we have:

Proposition 4.22. *Aperiodic monoids form a variety of finite monoids.*

We let **A** denote the variety of aperiodic monoids.

$\mathcal{J}$-trivial, $\mathcal{R}$-trivial and $\mathcal{L}$-trivial monoids

We let **J** [**R**, **L**], denote the variety of $\mathcal{J}$-trivial [$\mathcal{R}$-trivial, $\mathcal{L}$-trivial] monoids. The identities defining these varieties are given in the next proposition.

Proposition 4.23. *The following equalities hold*

$$\mathbf{R} = [\![(xy)^\omega x = (xy)^\omega]\!]$$
$$\mathbf{L} = [\![y(xy)^\omega = (xy)^\omega]\!]$$
$$\mathbf{J} = [\![y(xy)^\omega = (xy)^\omega = (xy)^\omega x]\!] = [\![x^{\omega+1} = x^\omega, (xy)^\omega = (yx)^\omega]\!]$$

Moreover, the identities $(x^\omega y^\omega)^\omega = (x^\omega y)^\omega = (xy^\omega)^\omega = (xy)^\omega$ *are satisfied by* **J**.

Proof. (1) Let M be a monoid and let $x, y \in M$. If ω is interpreted as the exponent of M, we observe that $(xy)^\omega x \mathrel{\mathcal{R}} (xy)^\omega$ since $((xy)^\omega x)(y(xy)^{\omega-1}) = (xy)^{2\omega} = (xy)^\omega$. Thus if M is $\mathcal{R}$-trivial, the identity $(xy)^\omega x = (xy)^\omega$ holds in M.

Conversely, assume that M satisfies the identity $(xy)^\omega x = (xy)^\omega$ and let u and v be two $\mathcal{R}$-equivalent elements of M. Then, there exist $x, y \in M$

such that $ux = v$ and $vy = u$. It follows that $u = uxy = u(xy)^\omega$ and thus $v = ux = u(xy)^\omega x$. Now, since $(xy)^\omega x = (xy)^\omega$, $u = v$ and M is $\mathcal{R}$-trivial.

(2) The proof is dual for the variety $\mathbf{L}$.

(3) Since $\mathbf{J} = \mathbf{R} \cap \mathbf{L}$, it follows from (1) and (2) that $\mathbf{J}$ is defined by the identities $y(xy)^\omega = (xy)^\omega = (xy)^\omega x$. Taking $y = 1$, we obtain $x^\omega = x^\omega x$ and also $(xy)^\omega = y(xy)^\omega = (yx)^\omega y = (yx)^\omega$. Conversely, suppose that a monoid satisfies the identities $x^{\omega+1} = x^\omega$ and $(xy)^\omega = (yx)^\omega$. Then we have $(xy)^\omega = (yx)^\omega = (yx)^{\omega+1} = y(xy)^\omega x$, whence $(xy)^\omega = y^\omega (xy)^\omega x^\omega = y^{\omega+1}(xy)^\omega x^\omega = y(xy)^\omega$ and likewise $(xy)^\omega = (xy)^\omega x$. □

Note that the following inclusions hold: $\mathbf{J} \subset \mathbf{R} \subset \mathbf{A}$ and $\mathbf{J} \subset \mathbf{L} \subset \mathbf{A}$.

Semilattices

A *semilattice* is an idempotent and commutative monoid. Semilattices form a variety, denoted by $\mathbf{J_1}$ and defined by the identities $x^2 = x$ and $xy = yx$.

Proposition 4.24. *The variety $\mathbf{J_1}$ is generated by the monoid U_1.*

Proof. Proposition 4.21 shows that the variety $\mathbf{J_1}$ is generated by its monogenic monoids. Now, there are only two monogenic idempotent monoids: the trivial monoid and the monoid U_1 considered in Section II.2.2. It follows that U_1 generates $\mathbf{J_1}$. □

Let us also mention a useful property

Proposition 4.25. *A $\mathcal{J}$-trivial idempotent monoid is a semilattice.*

Proof. Let M be a $\mathcal{J}$-trivial idempotent monoid. Proposition 4.23 show that M satisfies the identity $(xy)^\omega = (yx)^\omega$. Since M is idempotent, this identity becomes $xy = yx$ and thus M is commutative. Therefore M is a semilattice. □

Idempotent and $\mathcal{R}$-trivial [$\mathcal{L}$-trivial] monoids

Idempotent and $\mathcal{R}$-trivial [$\mathcal{L}$-trivial] monoids form a variety of monoids, denoted by $\mathbf{R_1}$ [$\mathbf{L_1}$].

Proposition 4.26. *The variety $\mathbf{R_1}$ [$\mathbf{L_1}$] is defined by the identity $xyx = xy$ [$xyx = yx$].*

Proof. We give only the proof for $\mathbf{R_1}$, since the case of $\mathbf{L_1}$ is symmetric.

Let M be an idempotent and $\mathcal{R}$-trivial monoid. Let $x, y \in M$. Since M is idempotent, one gets $xy = xyxy$ and thus $xy \mathrel{\mathcal{R}} xyx$. But M is $\mathcal{R}$-trivial and thus $xy = xyx$. It follows that $\mathbf{R_1}$ satisfies the identity $xy = xyx$.

Conversely, let M be a monoid satisfying the identity $xyx = xy$. Taking $y = 1$ gives $x^2 = x$ and thus M is idempotent. It follows also that M satisfies the identity $(xy)^\omega x = (xy)^\omega$ and hence is $\mathcal{R}$-trivial by Proposition 4.23. □

We shall see later (Proposition XIV.1.6) that the variety $\mathbf{R_1}$ [$\mathbf{L_1}$] is generated by the monoid U_2 [$\tilde{U}_2$].

The variety $\mathbb{D}\mathbf{S}$

A monoid belongs to the variety $\mathbb{D}\mathbf{S}$ if each of its regular $\mathcal{D}$-classes is a semigroup. In this case, every regular $\mathcal{D}$-class is completely regular (see Figure V.4.1).

The next proposition gathers various characterisations of $\mathbb{D}\mathbf{S}$ and shows in particular that $\mathbb{D}\mathbf{S} = [\![((xy)^\omega(yx)^\omega(xy)^\omega)^\omega = (xy)^\omega]\!]$.

Proposition 4.27. *Let M be a monoid. The following conditions are equivalent:*

(1) *M belongs to $\mathbb{D}\mathbf{S}$,*
(2) *M satisfies the identity $((xy)^\omega(yx)^\omega(xy)^\omega)^\omega = (xy)^\omega$,*
(3) *every regular $\mathcal{H}$-class of M is a group,*
(4) *if $s, t \in M$, s is regular and $s \leqslant_\mathcal{J} t$, then $s \mathcal{R} st$ and $s \mathcal{L} ts$,*
(5) *for each idempotent $e \in M$, the set*

$$M_e = \{s \in M \mid e \leqslant_\mathcal{J} s\}$$

is a subsemigroup of M.

Proof. (1) implies (2). Let $x, y \in M$. By Proposition V.2.22, $(xy)^\omega$ and $(yx)^\omega$ are conjugate idempotents. In particular, they belong to the same $\mathcal{D}$-class D. Since $M \in \mathbb{D}\mathbf{S}$, D is a simple semigroup and by Proposition 4.19, it satisfies the identity $(xyx)^\omega = x^\omega$. Therefore, condition (2) is satisfied.

(2) implies (3). Let x be a regular element, let D be its $\mathcal{D}$-class and let y be an inverse of x. Then $e = xy$ and $f = yx$ are two idempotents of D. The $\mathcal{H}$-class H of x is equal to $R(e) \cap L(f)$. Furthermore, condition (2) gives $(efe)^\omega = e$. It follows that $ef \mathcal{J} e \mathcal{J} fe$ and hence ef and fe also belong to D. Thus by Theorem V.1.11, $R(e) \cap L(f)$ contains an idempotent and Proposition V.1.13 shows that H is a group, which proves (3).

(3) implies (1). Let D be a regular $\mathcal{D}$-class. If each regular $\mathcal{H}$-class of D is a group, then each regular $\mathcal{H}$-class contains an idempotent and D is a semigroup by Theorem V.1.11.

Thus (1), (2) and (3) are equivalent. We now show that (1)–(3) implies (4). Let s be a regular element of M. By (3), $s \mathcal{H} (xty)^\omega = ((xty)^\omega(yxt)^\omega(xty)^\omega)^\omega$ and hence $(xty)^\omega \mathcal{J} t(xty)^\omega$. It follows that

$$s \mathcal{J} s^\omega = (xty)^\omega \mathcal{J} t(xty)^\omega = ts^\omega \leqslant_\mathcal{J} ts \leqslant_\mathcal{J} s$$

and hence $s \mathcal{J} ts$. It follows from Theorem V.1.9 that $s \mathcal{L} ts$. Similarly, $s \mathcal{R} st$.

(4) implies (3). Condition (4) applied with $s = t$ shows that, if s is regular, then $s \mathcal{R} s^2$ and $s \mathcal{L} s^2$. Therefore $s \mathcal{H} s^2$ and by Proposition V.1.13, the $\mathcal{H}$-class of s is a group, which establishes (3). Thus conditions (1)–(4) are equivalent.

(1)–(4) implies (5). Let e be an idempotent and let $s, t \in M_e$. Then $e \leqslant_\mathcal{J} s$, $e \leqslant_\mathcal{J} t$ and by (4), $te \mathcal{L} e$ and $e \mathcal{R} es$. Since $M \in \mathbb{D}\mathbf{S}$, the $\mathcal{D}$-class D of e is a semigroup. Since $es, te \in D$, one gets $este \in D$ and hence $e \mathcal{J} este \leqslant_\mathcal{J} st$. Thus $st \in M_e$. This proves (5).

(5) implies (1). Let D be a regular $\mathcal{D}$-class of M and let e be an idempotent of D. If $s, t \in D$, then $s, t \in M_e$ and by (5), $st \in M_e$, that is $e \leqslant_\mathcal{J} st$. Since $st \leqslant_\mathcal{J} s \mathcal{J} e$, one has $e \mathcal{J} st$ and hence $st \in D$. □

It is also useful to characterise the monoids that are not in $\mathbb{D}\mathbf{S}$.

Proposition 4.28. *Let M be a monoid. The following conditions are equivalent:*

(1) *M does not belong to $\mathbb{D}\mathbf{S}$,*

(2) *there exist two idempotents $e, f \in M$ such that $e \mathcal{J} f$ but $ef \not\mathcal{J} e$,*

(3) *B_2^1 divides $M \times M$.*

Proof. (1) implies (2). If M does not belong to $\mathbb{D}\mathbf{S}$, M contains a regular $\mathcal{D}$-class D which is not a semigroup. By Proposition V.4.36, one can find two idempotent $e, f \in D$ such that $ef \notin D$.

(2) implies (3). Let D be the $\mathcal{D}$-class of e. Since $e \mathcal{J} f$, Proposition V.2.22 shows that there exist two elements $s, \bar{s} \in D$ such that $s\bar{s} = e$ and $\bar{s}s = f$. Since $ef \notin D$, Theorem V.1.11 shows that $L(e) \cap R(f)$ contains no idempotent and thus $\bar{s}^2 \notin D$. Let N be the submonoid of $M \times M$ generated by the elements $a = (s, \bar{s})$ and $\bar{a} = (\bar{s}, s)$. Then a and $\bar{a}$ are mutually inverse in N and the element $a\bar{a} = (e, f)$ and $\bar{a}a = (f, e)$ are idempotent. Therefore, the four elements a, $\bar{a}$, $a\bar{a}$ and $\bar{a}a$ form a regular $\mathcal{D}$-class C.

a	$^*a\bar{a}$
$^*\bar{a}a$	$\bar{a}$

Furthermore, $aa = (ss, \bar{s}\bar{s})$ and $bb = (\bar{s}\bar{s}, ss)$. Since $\bar{s}^2 \notin D$, it follows that aa and bb are not in C and in fact $N - (C \cup \{1\})$ is the ideal of N consisting of all elements x such that $x <_{\mathcal{J}} a$. It follows that N/J is isomorphic to B_2^1, which proves (2).

(3) implies (1). Suppose that $M \in \mathbb{D}\mathbf{S}$. Since $\mathbb{D}\mathbf{S}$ is a variety of monoids, $M \times M \in \mathbb{D}\mathbf{S}$ and since B_1^2 divides $M \times M$, B_1^2 also belongs to $\mathbb{D}\mathbf{S}$, a contradiction. □

The variety $\mathbb{D}\mathbf{A}$

A monoid M belongs to the variety $\mathbb{D}\mathbf{A}$ if each regular $\mathcal{D}$-class of M is an aperiodic semigroup. It is equivalent to require that every regular $\mathcal{D}$-class is a rectangular band.

$*$	$*$		$\cdots$		$*$
$*$	$*$		$\cdots$		$*$
$\vdots$	$\vdots$		$\ddots$		$\vdots$
$*$	$*$		$\cdots$		$*$

Figure 4.2. A regular $\mathcal{D}$-class in a monoid of $\mathbb{D}\mathbf{A}$.

The next proposition gathers various characterisations of $\mathbb{D}\mathbf{A}$ and shows in particular that $\mathbb{D}\mathbf{A} = \mathbb{D}\mathbf{S} \cap \mathbf{A} = [\![(xy)^\omega(yx)^\omega(xy)^\omega = (xy)^\omega, x^{\omega+1} = x^\omega]\!]$.

Proposition 4.29. *Let M be a monoid. The following conditions are equivalent:*

(1) *M belongs to $\mathbb{D}\mathbf{A}$,*
(2) *M is aperiodic and belongs to $\mathbb{D}\mathbf{S}$,*
(3) *M satisfies the identities $(xy)^\omega(yx)^\omega(xy)^\omega = (xy)^\omega$ and $x^{\omega+1} = x^\omega$,*
(4) *for each $e \in E(M)$ and each $s \in M$, $e \leqslant_{\mathcal{J}} s$ implies $ese = e$,*
(5) *M is aperiodic and for each $e, f \in E(M)$, $e \mathcal{J} f$ implies $efe = e$.*

Proof. We prove (3) ⇒ (2) ⇒ (1) ⇒ (4) ⇒ (5) ⇒ (3) in this order.

(3) implies (2) follows from Proposition 4.27.

(2) implies (1). If $M \in \mathbb{D}\mathbf{S}$, each regular $\mathcal{D}$-class is completely regular and hence is a simple semigroup. Furthermore, if M is aperiodic, each regular $\mathcal{D}$-class is a simple aperiodic semigroup and hence is idempotent by Corollary V.3.34.

(1) implies (4). Let $M \in \mathbb{D}\mathbf{A}$. Let $e \in E(M)$ and $s \in M$ be such that $e \leqslant_{\mathcal{J}} s$. Since $M \in \mathbb{D}\mathbf{S}$, Proposition 4.27 (5) shows that the set $M_e = \{x \in M \mid e \leqslant_{\mathcal{J}} x\}$ is a subsemigroup of M. Since s and e belong to M_e, one also has $ese \in M_e$, and hence $e \leqslant_{\mathcal{J}} ese$. Since $ese \leqslant_{\mathcal{J}} e$ one gets finally $ese \mathcal{J} e$. But each regular $\mathcal{D}$-class of M is a rectangular band, and thus $ese = e$.

(4) implies (5). For each $x \in M$, one has $x^\omega \leqslant_{\mathcal{J}} x$, and thus by (4), $x^{\omega+1} = x^\omega$. Therefore M is aperiodic. Moreover if e and f are two $\mathcal{J}$-equivalent idempotents of M, then $e \leqslant_{\mathcal{J}} f$ and by (4), $efe = e$.

(5) implies (3). Suppose that M satisfies (5). Then M is aperiodic and satisfies the identity $x^{\omega+1} = x^\omega$. Furthermore if $x, y \in M$, the elements $(xy)^\omega$ and $(yx)^\omega$ are two conjugate idempotents, which are $\mathcal{J}$-equivalent by Proposition V.2.22. It follows that $(xy)^\omega(yx)^\omega(xy)^\omega = (xy)^\omega$. □

4.3 Varieties of ordered monoids

We denote by $\mathbf{N}^+$ [$\mathbf{N}^-$] the variety of ordered semigroups defined by the identities $yx^\omega = x^\omega = x^\omega y$ and $y \leqslant x^\omega$ [$yx^\omega = x^\omega = x^\omega y$ and $x^\omega \leqslant y$].

We let $\mathbf{J}_1^+$ [$\mathbf{J}_1^-$] denote the variety of ordered monoids defined by the identities $x^2 = x$, $xy = yx$ and $1 \leqslant x$ [$x \leqslant 1$]. Proposition 4.24 can be readily adapted to the ordered case.

Proposition 4.30. *The variety $\mathbf{J}_1^+$ [$\mathbf{J}_1^-$] is generated by the monoid U_1^+ [U_1^-].*

We let $\mathbf{J}^+$ [$\mathbf{J}^-$] denote the variety of ordered monoids defined by the identity $x \leqslant 1$ [$1 \leqslant x$]. The notation is justified by Proposition VII.3.12, which states that every finite ordered monoid satisfying the identity $1 \leqslant x$ is $\mathcal{J}$-trivial.

A locally positive $\mathcal{J}$-trivial semigroup is an ordered semigroup S such that, for each idempotent $e \in S$, the ordered semigroup eSe is positive $\mathcal{J}$-trivial. Locally positive $\mathcal{J}$-trivial semigroups form a variety of ordered semigroups, denoted by $\mathbb{L}\mathbf{J}^+$ and defined by the identity $[\![x^\omega \leqslant x^\omega y x^\omega]\!]$. This variety plays a crucial role in the study of the polynomial closure (see Chapter XV).

4.4 Summary

We summarise in the next table the varieties defined so far.

Notation	Name	Profinite identities
G	Groups	$[\![x^\omega = 1]\!]$
Com	Commutative monoids	$[\![xy = yx]\!]$
J	$\mathcal{J}$-trivial monoids	$[\![y(xy)^\omega = (xy)^\omega = (xy)^\omega x]\!]$
R	$\mathcal{R}$-trivial monoids	$[\![(xy)^\omega x = (xy)^\omega]\!]$
L	$\mathcal{L}$-trivial monoids	$[\![y(xy)^\omega = (xy)^\omega]\!]$
$\mathbf{J_1}$	Semilattices	$[\![x^2 = x, xy = yx]\!]$
$\mathbf{R_1}$	Idempotent and $\mathcal{R}$-trivial	$[\![xyx = xy]\!]$
$\mathbf{L_1}$	Idempotent and $\mathcal{L}$-trivial	$[\![xyx = yx]\!]$
A	Aperiodic monoids	$[\![x^{\omega+1} = x^\omega]\!]$
$\mathbb{D}\mathbf{A}$	Regular $\mathcal{D}$-classes are aperiodic semigroups	$[\![(xy)^\omega(yx)^\omega(xy)^\omega = (xy)^\omega, x^{\omega+1} = x^\omega]\!]$
$\mathbb{D}\mathbf{S}$	Regular $\mathcal{D}$-classes are semigroups	$[\![((xy)^\omega(yx)^\omega(xy)^\omega)^\omega = (xy)^\omega]\!]$

Notation	Name	Profinite identities
N	Nilpotent semigroups	$[\![yx^\omega = x^\omega = x^\omega y]\!]$
$\ell\mathbf{1}$ or **K**	Lefty trivial semigroups	$[\![x^\omega y = x^\omega]\!]$
r**1** or **D**	Righty trivial semigroups	$[\![yx^\omega = x^\omega]\!]$
$\mathbb{L}\mathbf{1}$	Locally trivial semigroups	$[\![x^\omega y x^\omega = x^\omega]\!]$
$\mathbb{L}\mathbf{G}$	Locally groups	$[\![(x^\omega y x^\omega)^\omega = x^\omega]\!]$
CS	Simple semigroups	$[\![x^{\omega+1} = x, (xyx)^\omega = x^\omega]\!]$
$\mathbf{N}^+$		$[\![yx^\omega = x^\omega = x^\omega y, y \leqslant x^\omega]\!]$
$\mathbf{N}^-$		$[\![yx^\omega = x^\omega = x^\omega y, x^\omega \leqslant y]\!]$
$\mathbf{J_1^+}$		$[\![x^2 = x, xy = yx, 1 \leqslant x]\!]$
$\mathbf{J_1^-}$		$[\![x^2 = x, xy = yx, x \leqslant 1]\!]$
$\mathbf{J}^+$		$[\![1 \leqslant x]\!]$
$\mathbf{J}^-$		$[\![x \leqslant 1]\!]$
$\mathbb{L}\mathbf{J}^+$		$[\![x^\omega \leqslant x^\omega y x^\omega]\!]$

5 Exercises

Section 1

In Exercises 1 and 2, we exceptionally consider infinite monoids.

Exercise 1. A *Birkhoff variety* of monoids is a class $\mathbf{V}$ of (possibly infinite) monoids such that:

(1) if $S \in \mathbf{V}$ and if T is a submonoid of S, then $S \in \mathbf{V}$,

(2) if $S \in \mathbf{V}$ and if T is a quotient of S, then $S \in \mathbf{V}$,
(3) if $(S_i)_{i\in I}$ is a (possibly infinite) family of monoids of $\mathbf{V}$, the product $\prod_{i\in I} S_i$ is also in $\mathbf{V}$.

Show that the class of all commutative monoids forms a Birkhoff variety, but that the class of all groups does not (consider the submonoid $\mathbb{N}$ of Z).

Exercise 2. Let A be an alphabet and let $u, v \in A^*$. A monoid M *satisfies the identity* $u = v$ if and only if, for each monoid morphism $\varphi : A^* \to M$, $\varphi(u) = \varphi(v)$.
Show that a class of monoids is a Birkhoff variety if and only if it can be defined by a set of identities (Birkhoff's theorem). Birkhoff's theorem can be extended to any variety of algebras, including semigroups, ordered semigroups and ordered monoids.

Section 2
We are now back with the usual convention of this chapter: semigroups are either finite or free.

Exercise 3. Let $\mathbf{G}$ denote the variety of finite groups and let $d_{\mathbf{G}}$ be the function defined on page 190. Thus one has

$$r_{\mathbf{G}}(u, v) = \min\{\mathrm{Card}(G) \mid G \text{ is a finite group separating } u \text{ and } v\}$$

and $d_{\mathbf{G}}(u, v) = 2^{-r_{\mathbf{G}}(u,v)}$ with the usual convention $\min \emptyset = +\infty$ and $2^{-\infty} = 0$.

(1) Given two distinct words, prove that there exists a permutation automaton accepting one of the words and rejecting the other one. Conclude that $d_{\mathbf{G}}$ is a metric on A^*.
(2) We are interested now in the topology on A^* defined by $d_{\mathbf{G}}$. Show that for each $u \in A^*$, one has $\lim_{n\to\infty} u^{n!} = 1$. Conclude that if L is a closed subset of A^*, and if $xu^+y \subseteq L$, the $xu^*y \subseteq L$.
(3) Show that the group languages of A^* are clopen in the metric space $(A^*, d_{\mathbf{G}})$.

Section 3

Exercise 4. Let $\mathbf{G_S}$ be the variety of *semigroups* generated by all groups. Show that $\mathbf{G_S} = [\![x^\omega y = yx^\omega = y]\!]$.

Exercise 5. Show that the variety $\mathbf{N}$ is defined by the identity $yx^\omega z = x^\omega$. Show that the group $\mathbb{Z}/2\mathbb{Z}$ satisfies the semigroup identity $yx^\omega y = x^\omega$. Conclude that this identity does not define $\mathbf{N}$.

Section 4

Exercise 6. Let $\mathbf{V}$ be a variety of [ordered] monoids. A semigroup S is *locally* $\mathbf{V}$ if, for every idempotent e of S, the [ordered] semigroup eSe belongs to $\mathbf{V}$. Show that the locally $\mathbf{V}$ [ordered] semigroups form a variety of [ordered] semigroups, denoted by $\mathbb{L}\mathbf{V}$.

Exercise 7. Let $\mathbf{V}$ be the variety of monoids. Given a monoid M, let $\langle E(M)\rangle$ denote the submonoid of M generated by the idempotents of M. Show that the monoids M such that $\langle E(M)\rangle$ belongs to $\mathbf{V}$ form a variety of monoids, denoted by $\mathbb{E}\mathbf{V}$.

Exercise 8. Let $\mathbf{V}$ be the class of all finite aperiodic monoids M such that $es = se$ for all $e \in E(M)$ and all $s \in M$.

(1) Show that $\mathbf{V}$ is a variety of finite monoids.
(2) Show that every monoid of $\mathbf{V}$ has a zero.
(3) Let M be a monoid of $\mathbf{V}$ and let $e \in E(M)$. Show that the function $\alpha : M \to eM$ defined by $\alpha(s) = es$ is a monoid morphism.
(4) Let π be the projection from M onto M/MeM. Show that the function $\gamma : M \to eM \times (M/MeM)$ defined by $\gamma(s) = (\alpha(s), \pi(s))$ is an injective morphism.
(5) Deduce that $\mathbf{V}$ is generated by the monoids of the form S^1, where S is a nilpotent semigroup.

6 Notes

Reference books on varieties include [4, 42, 98]. Varieties of ordered monoids were first introduced in [101].

Chapter XII

Equations and languages

In this chapter, we show how profinite words can be used to give algebraic descriptions of certain classes of recognisable languages.

1 Equations

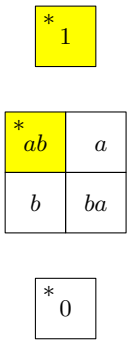
Formally, a *profinite equation* is a pair (u, v) of profinite words of $\widehat{A^*}$. We also use the term *explicit equation* when both u and v are words of A^*. We say that a recognisable language L of A^* *satisfies the profinite equation* $u \to v$ (or $v \leftarrow u$) if the condition $u \in \overline{L}$ implies $v \in \overline{L}$.

This formal definition somewhat obscures the intuitive meaning of this notion, but fortunately Proposition X.3.17 provides some more practical criteria:

Corollary 1.1. *Let L be a recognisable language of A^*, let η be its syntactic morphism and let φ be any morphism onto a finite monoid recognising L. The following conditions are equivalent:*

(1) *L satisfies the equation $u \to v$,*
(2) *$\widehat{\eta}(u) \in \eta(L)$ implies $\widehat{\eta}(v) \in \eta(L)$,*
(3) *$\widehat{\varphi}(u) \in \varphi(L)$ implies $\widehat{\varphi}(v) \in \varphi(L)$.*

Example 1.1. Consider the language $L = (ab)^*$ on the alphabet $A = \{a, b\}$. It is not difficult to see that its syntactic monoid is the monoid $B_2^1 = \{1, a, b, ab, ba, 0\}$ defined by the presentation $aba = a$, $bab = b$, $aa = bb = 0$ (see Example IV.3.4). The syntactic morphism $\eta : A^* \to B_2^1$ is defined by $\eta(a) = a$ and $\eta(b) = b$ and the syntactic image $P = \eta(L)$ is equal to $\{1, ab\}$.



Let (u, v) be a pair of profinite words of $\widehat{A^*}$. Then L satisfies the equation $u \to v$ if and only if $\widehat{\eta}(u) \in P$ implies $\widehat{\eta}(v) \in P$. Since in logic, "p implies q" is equivalent to "q or not p", L satisfies $u \to v$ if and only if $\widehat{\eta}(v) \in P$ or $\widehat{\eta}(u) \notin P$. For instance, since $\widehat{\eta}(aa) = 0 \notin P$, L satisfies the equations $aa \to v$ for all profinite words v. In the same way, since $\widehat{\eta}((ab)^\omega) = ab \in P$, L satisfies the equations $u \to (ab)^\omega$ for all profinite words u.

The following result shows that equations behave very nicely with respect to complement.

Proposition 1.2. *Let L be a recognisable language of A^*. Then L satisfies the equation $u \to v$ if and only if L^c satisfies the equation $v \to u$.*

Proof. Let η be the syntactic morphism of L, which is also the syntactic morphism of L^c. By Corollary 1.1, L satisfies the equation $u \to v$ if and only if

$$\widehat{\eta}(u) \in \eta(L) \text{ implies } \widehat{\eta}(v) \in \eta(L) \tag{1.1}$$

Similarly, L^c satisfies the equation $v \to u$ if and only if $\widehat{\eta}(v) \in \eta(L^c)$ implies $\widehat{\eta}(u) \in \eta(L^c)$. But since $\eta(L^c) = \eta(L)^c$, the later implication can be written as $\widehat{\eta}(v) \notin \eta(L)$ implies $\widehat{\eta}(u) \notin \eta(L)$, which is the contrapositive of (1.1). □

2 Equational characterisation of lattices

Given a set E of equations of the form $u \to v$, the subset of $\mathrm{Rec}(A^*)$ *defined by* E is the set of all recognisable languages of A^* satisfying all the equations of E.

A *lattice of languages* of A^* is a set of languages of A^* containing the empty language $\emptyset$, the full language A^* and which is closed under finite union and finite intersection. A lattice of languages closed under complementation is a *Boolean algebra of languages.*

Proposition 2.3. *The set of recognisable languages of A^* defined by a set of equations forms a lattice of languages.*

Proof. Let E be a set of equations and let $\mathcal{L}$ be the class of languages of A^* defined by E. We claim that $\mathcal{L}$ is a lattice. First, it is clear that the empty language and the full language A^* satisfy any equation of the form $u \to v$. Therefore, these two languages are in $\mathcal{L}$.

Let now L_1 and L_2 be languages of $\mathcal{L}$ and let $u \to v$ be an equation of E. Then, for $i = 1, 2$, the condition $u \in \overline{L}_i$ implies $v \in \overline{L}_i$. Recall that, by Theorem X.3.19, one has $\overline{L_1 \cup L_2} = \overline{L}_1 \cup \overline{L}_2$ and $\overline{L_1 \cap L_2} = \overline{L}_1 \cap \overline{L}_2$. Suppose that $u \in \overline{L_1 \cup L_2}$. Then either $u \in \overline{L}_1$ or $u \in \overline{L}_2$ and thus either $v \in \overline{L}_1$ or $v \in \overline{L}_2$ and finally $v \in \overline{L_1 \cup L_2}$. Similarly, if $u \in \overline{L_1 \cap L_2}$, then $u \in \overline{L}_1$ and $u \in \overline{L}_2$, whence $v \in \overline{L}_1$ and $v \in \overline{L}_2$ and finally $v \in \overline{L_1 \cap L_2}$.

It follows that $L_1 \cap L_2$ and $L_1 \cup L_2$ satisfy all the equations of E and thus $\mathcal{L}$ is a lattice of languages. □

Our aim is now to show that the converse of Proposition 2.3 also holds. We start with a result on languages interesting on its own right.

Proposition 2.4. *Let L, $L_1, \ldots, L_n$ be recognisable languages. If L satisfies all the explicit equations satisfied by $L_1, \ldots, L_n$, then L belongs to the lattice of languages generated by $L_1, \ldots, L_n$.*

Proof. We claim that

$$L = \bigcup_{I \in \mathcal{I}} \bigcap_{i \in I} L_i \tag{2.1}$$

where $\mathcal{I}$ is the set of all subsets I of $\{1, \ldots, n\}$ for which there exists a word $v \in L$ such that $v \in L_i$ if and only if $i \in I$.

Let R be the right-hand side of (2.1). If $u \in L$, let $I = \{i \mid u \in L_i\}$. By construction, $I \in \mathcal{I}$ and $u \in \cap_{i \in I} L_i$. Thus $u \in R$. This proves the inclusion $L \subseteq R$.

To prove the opposite direction, consider a word $u \in R$. By definition, there exists a set $I \in \mathcal{I}$ such that $u \in \cap_{i \in I} L_i$ and a word $v \in L$ such that $v \in L_i$ if and only if $i \in I$. We claim that the equation $v \to u$ is satisfied by each language L_i. Indeed, if $i \in I$, then $u \in L_i$ by definition. If $i \notin I$, then $v \notin L_i$ by definition of I, which proves the claim. It follows that $v \to u$ is also satisfied by L. Since $v \in L$, it follows that $u \in L$. This concludes the proof of (2.1) and shows that L belongs to the lattice of languages generated by $L_1, \ldots, L_n$. □

An important consequence of Proposition 2.4 is that finite lattices of recognisable languages can be defined by explicit equations.

Corollary 2.5. *A finite set of recognisable languages of A^* is a lattice of recognisable languages if and only if it can be defined by a set of explicit equations of the form $u \to v$, where $u, v \in A^*$.*

Proof. Proposition 2.3 shows that a set of recognisable languages defined by a set of equations form a lattice of languages.

To prove the opposite direction, consider a finite lattice $\mathcal{L}$ of recognisable languages and let E be the set of explicit equations satisfied by all the languages of $\mathcal{L}$. Proposition 2.4 shows that any language L that satisfies the equations of E belongs to $\mathcal{L}$. Thus $\mathcal{L}$ is defined by E. □

We now are ready for the main result.

Theorem 2.6. *A set of recognisable languages of A^* is a lattice of recognisable languages if and only if it can be defined by a set of equations of the form $u \to v$, where $u, v \in \widehat{A^*}$.*

Proof. For each recognisable language L, set

$$E_L = \{(u, v) \in \widehat{A^*} \times \widehat{A^*} \mid L \text{ satisfies } u \to v\}$$

Lemma 2.7. *For each recognisable language L, E_L is a clopen subset of $\widehat{A^*} \times \widehat{A^*}$.*

Proof. One has

$$\begin{aligned} E_L &= \{(u, v) \in \widehat{A^*} \times \widehat{A^*} \mid L \text{ satisfies } u \to v\} \\ &= \{(u, v) \in \widehat{A^*} \times \widehat{A^*} \mid u \in \overline{L} \text{ implies } v \in \overline{L}\} \\ &= \{(u, v) \in \widehat{A^*} \times \widehat{A^*} \mid v \in \overline{L} \text{ or } u \notin \overline{L}\} \\ &= (\overline{L}^c \times \widehat{A^*}) \cup (\widehat{A^*} \times \overline{L}) \end{aligned}$$

The result follows since, by Proposition X.3.16, $\overline{L}$ is clopen. □

Let $\mathcal{L}$ be a lattice of recognisable languages and let E be the set of profinite equations satisfied by all languages of $\mathcal{L}$. We claim that E defines $\mathcal{L}$. First, by definition, every language of $\mathcal{L}$ satisfies the equations of E. It just remains to prove that if a language L satisfies the equations of E, then L belongs to $\mathcal{L}$.

First observe that the set

$$\{E_L\} \cup \{E_K^c \mid K \in \mathcal{L}\}$$

is a covering of $\widehat{A^*} \times \widehat{A^*}$. Indeed, if $(u, v) \notin \cup_{K\in\mathcal{L}} E_K^c$, then $(u, v) \in \cap_{K\in\mathcal{L}} E_K$, which means that $u \to v$ is an equation satisfied by all languages of $\mathcal{L}$. It follows that L also satisfies this equation, and thus $(u, v) \in E_L$. Furthermore, Lemma 2.7 shows that the elements of this covering are open sets. Since $\widehat{A^*} \times \widehat{A^*}$ is compact, it admits a finite subcovering, and we may assume that this covering contains E_L and is equal to

$$\{E_L\} \cup \{E_{L_1}^c, \ldots, E_{L_n}^c\}$$

for some languages $L_1, \ldots, L_n$ of $\mathcal{L}$. By the same argument as above, it follows that if an equation $u \to v$ is satisfied by $L_1, \ldots, L_n$, then it is satisfied by L. By Proposition 2.4, L belongs to the lattice of languages generated by $L_1, \ldots, L_n$ and hence belongs to $\mathcal{L}$. □

Writing $u \leftrightarrow v$ for ($u \to v$ and $v \to u$), we get an equational description of the Boolean algebras of recognisable languages.

Corollary 2.8. *A set of recognisable languages of A^* is a Boolean algebra of recognisable languages if and only if it can be defined by a set of equations of the form $u \leftrightarrow v$, where $u, v \in \widehat{A^*}$.*

We now specialise Theorem 2.6 and Corollary 2.8 to lattices of closed under quotient. Further specialisations to streams and varieties of languages will form the topic of the next chapter.

3 Lattices of languages closed under quotients

We say that a class $\mathcal{L}$ of languages is *closed under quotients* if for every $L \in \mathcal{L}$ and $u \in A^*$, $u^{-1}L$ and Lu^{-1} are also in $\mathcal{L}$.

Let u and v be two profinite words of $\widehat{A^*}$. We say that a recognisable language L *satisfies the equation* $u \leqslant v$ if, for all $x, y \in \widehat{A^*}$, it satisfies the equation $xuy \to xvy$. Since A^* is dense in $\widehat{A^*}$, it is equivalent to state that L satisfies these equations only for all $x, y \in A^*$. But there is a much more convenient characterisation using the syntactic ordered monoid of L.

Proposition 3.9. *Let L be a recognisable language of A^*, let $(M, \leqslant_L)$ be its syntactic ordered monoid and let $\eta : A^* \to M$ be its syntactic morphism. Then L satisfies the equation $u \leqslant v$ if and only if $\widehat{\eta}(u) \leqslant_L \widehat{\eta}(v)$.*

Proof. Corollary 1.1 shows that L satisfies the equation $u \leqslant v$ if and only if, for every $x, y \in A^*$, $\widehat{\eta}(xvy) \in \eta(L)$ implies $\widehat{\eta}(xvy) \in \eta(L)$. Since $\widehat{\eta}(xuy) = \widehat{\eta}(x)\widehat{\eta}(u)\widehat{\eta}(y) = \eta(x)\widehat{\eta}(u)\eta(y)$ and since η is surjective, this is equivalent to saying that, for all $s, t \in M$, $s\widehat{\eta}(u)t \in \eta(L)$ implies $s\widehat{\eta}(v)t \in \eta(L)$, which exactly means that $\widehat{\eta}(u) \leqslant_L \widehat{\eta}(v)$. □

We can now state the equational characterisation of lattices of recognisable languages closed under quotients.

Theorem 3.10. *A set of recognisable languages of A^* is a lattice of languages closed under quotients if and only if it can be defined by a set of equations of the form $u \leqslant v$, where $u, v \in \widehat{A^*}$.*

Proof. Let L be a recognisable language satisfying the equation $u \leqslant v$ and let $x, y \in A^*$. Since L satisfies the equation $xuy \to xvy$, the condition $xuy \in \overline{L}$ implies $xvy \in \overline{L}$ and hence $u \in x^{-1}\overline{L}y^{-1}$ implies $v \in x^{-1}\overline{L}y^{-1}$. By Corollary X.3.22, $x^{-1}\overline{L}y^{-1} = \overline{x^{-1}Ly^{-1}}$. Therefore $u \in \overline{x^{-1}Ly^{-1}}$ implies $v \in \overline{x^{-1}Ly^{-1}}$ and hence the language $x^{-1}Ly^{-1}$ satisfies the equation $u \leqslant v$. It follows that the set of recognisable languages defined by a set of equations of the form $u \leqslant v$ is a lattice of recognisable languages closed under quotients.

In the opposite direction, let $\mathcal{L}$ be a lattice of recognisable languages of A^* closed under quotients. By Theorem 2.6, $\mathcal{L}$ can be defined by a set E of equations of the form $u \to v$, where $u, v \in \widehat{A^*}$. Let now $u \to v$ be an equation of E, L a language of $\mathcal{L}$ and x, y two words of A^*. Since $\mathcal{L}$ is closed under quotient, $x^{-1}Ly^{-1}$ belongs to $\mathcal{L}$ and thus satisfies also the equation $u \to v$. It follows that L satisfies the equation $xuy \to xvy$ and hence L satisfies the equation $u \leqslant v$. It follows that $\mathcal{L}$ is defined by the equations of the form $u \leqslant v$ where $u \to v$ is an equation of E. □

Theorem 3.10 can be readily extended to Boolean algebras. Let u and v be two profinite words. We say that a recognisable language L *satisfies the equation* $u = v$ if it satisfies the equations $u \leqslant v$ and $v \leqslant u$. Proposition 3.9 now gives immediately:

Proposition 3.11. *Let L be a recognisable language of A^* and let η be its syntactic morphism. Then L satisfies the equation $u = v$ if and only if $\widehat{\eta}(u) = \widehat{\eta}(v)$.*

This leads to the following equational description of Boolean algebras of recognisable languages closed under quotients.

Corollary 3.12. *A set of recognisable languages of A^* is a Boolean algebra of languages closed under quotients if and only if it can be defined by a set of equations of the form $u = v$, where $u, v \in \widehat{A^*}$.*

4 Equational descriptions of lattices of languages

We first give a number of examples of lattices of languages. We start by revisiting some classical examples studied between 1960 and 1980. Then we consider some more recent examples.

4.1 The role of the zero

An element 0 of a monoid M is a *zero* if, for all $m \in M$, $0m = 0 = 0m$. It is easy to see that a monoid has at most one zero element. This allows one to use the notation 0 for the zero without ambiguity. Observe that the trivial monoid has a zero, but this is the only case for which $1 = 0$. Also note that the minimal ideal of a monoid with zero reduces to $\{0\}$. A quotient of a monoid with zero also has a zero, but this is not the case for a submonoid.

Languages with zero

A *language with zero* is a language whose syntactic monoid has a zero, or equivalently, a language fully recognised by a monoid with zero.

Proposition 4.13. *The class of recognisable languages with zero is closed under Boolean operations and under quotients. It is not closed under inverses of morphisms, even for length-preserving morphisms.*

Proof. Let $\mathcal{C}$ be the class of recognisable languages with zero. The empty language and the full language belong to $\mathcal{C}$, since their syntactic monoid is trivial. Proposition IV.2.9 shows that $\mathcal{C}$ is closed under complement. Let L_1 and L_2 be two languages of $\mathcal{C}$ and let $\varphi_1 : A^* \to M_1$ and $\varphi_2 : A^* \to M_2$ be their respective syntactic morphisms. By Proposition IV.2.10, $L_1 \cap L_2$ is recognised by the morphism $\varphi : A^* \to \mathrm{Im}(\varphi) \subseteq M_1 \times M_2$ defined by $\varphi(u) = (\varphi_1(u_1), \varphi_2(u_2))$. We claim that the monoid $M = \mathrm{Im}(\varphi)$ has a zero. Indeed, let u_1 and u_2 be words such that $\varphi_1(u_1) = 0$ and $\varphi_2(u_2) = 0$. Then the element $\varphi(u_1u_2)$ is a zero of M. Indeed, if $\varphi(u)$ is an element of M, then $\varphi(u)\varphi(u_1u_2) = \varphi(uu_1u_2) = (\varphi_1(uu_1u_2), \varphi_2(uu_1u_2)) = (0,0) = \varphi(u_1u_2)$ and similarly, $\varphi(u_1u_2)\varphi(u) = \varphi(u_1u_2)$. This proves the claim and shows that $L_1 \cap L_2$ is a language with zero. Thus $\mathcal{C}$ is closed under Boolean operations.

Let L be a recognisable language with zero and let M be its syntactic monoid. Let u be a word of A^*. It follows from Proposition IV.2.12 that the languages $u^{-1}L$ and Lu^{-1} are also recognised by M. It follows that the syntactic monoid of these languages is a quotient of M and hence has a zero. Thus $\mathcal{C}$ is closed under quotients.

Finally let $A = \{a, b\}$ and let $\varphi : A^* \to A^*$ be the length-preserving morphism defined by $\varphi(a) = \varphi(b) = a$. If $L = (a^2)^*$, then $\varphi^{-1}(L) = (A^2)^*$. Now L has a zero, but the syntactic monoid of $(A^2)^*$ is the cyclic group of order 2, which has no zero. □

According to Corollary 3.12, the class of recognisable languages with zero has an equational definition, but finding one explicitly requires a little bit of work.

Let us fix a total order on the alphabet A. Let $u_0, u_1, \ldots$ be the ordered sequence of all words of A^+ in the induced shortlex order. For instance, if $A = \{a, b\}$ with $a < b$, the first elements of this sequence would be

$$1, a, b, aa, ab, ba, bb, aaa, aab, aba, abb, baa, bab, bba, bbb, aaaa, \ldots$$

It is proved in [132, 6] that the sequence of words $(v_n)_{n \geqslant 0}$ defined by

$$v_0 = u_0, \ v_{n+1} = (v_n u_{n+1} v_n)^{(n+1)!}$$

converges to an idempotent ρ_A of the minimal ideal of the free profinite monoid on A. This profinite element can be used to give an equational characterisation of the recognisable languages with zero.

Proposition 4.14. *A recognisable language has a zero if and only if it satisfies the equation $x\rho_A = \rho_A = \rho_A x$ for all $x \in A^*$.*

Proof. Let L be a recognisable language and let $\eta : A^* \to M$ be its syntactic monoid. Since ρ_A belongs to the minimal ideal of A^*, $\eta(\rho_A)$ is an element of the

minimal ideal of M. In particular, if M has a zero, $\eta(\rho_A) = 0$ and L satisfies the equations $x\rho_A = \rho_A = \rho_A x$ for all $x \in A^*$.

Conversely, assume that L satisfies these equations. Let $m \in M$ and let $x \in A^*$ be such that $\eta(x) = m$. Then the equations $\eta(x\rho_A) = \eta(\rho_A) = \eta(\rho_A x)$ give $m\eta(\rho_A) = \eta(\rho_A) = \eta(\rho_A)m$, showing that $\eta(\rho_A)$ is a zero of M. Thus L has a zero. □

In the sequel, we shall use freely the symbol 0 in equations to mean that a monoid has a zero. For instance the equation $0 \leqslant x$ of Theorem 4.17 below should be formally replaced by the two equations $x\rho_A = \rho_A = \rho_A x$ and $\rho_A \leqslant x$.

Nondense languages

A language L of A^* is *dense* if, for every word $u \in A^*$, $L \cap A^*uA^* \neq \emptyset$, or, equivalently, if $(A^*)^{-1}L(A^*)^{-1} = A^*$. Note that dense languages are not closed under intersection: $(A^2)^*$ and $(A^2)^*A \cup \{1\}$ are dense, but their intersection is not dense.

Proposition 4.15. *Nondense or full languages are closed under finite union, finite intersection and left and right quotients.*

Proof. Let L_1 and L_2 be two nondense languages of A^*. Then there exist two words $u_1, u_2 \in A^*$ such that $L_1 \cap A^*u_1A^* = \emptyset$ and $L_2 \cap A^*u_2A^* = \emptyset$. It follows that $(L_1 \cap L_2) \cap A^*u_1A^* = \emptyset$ and $(L_1 \cup L_2) \cap A^*u_1u_2A^* = \emptyset$. Thus $L_1 \cap L_2$ and $L_1 \cup L_2$ are nondense. If $L_1 = A^*$, then $L_1 \cap L_2 = L_2$ and $L_1 \cup L_2 = A^*$. Thus nondense or full languages are closed under finite union and finite intersection.

Let L be a nondense language. Then there exists a word $u \in A^*$ such that $L \cap A^*uA^* = \emptyset$. Let $x, y \in A^*$. We claim that $x^{-1}Ly^{-1} \cap A^*uA^* = \emptyset$. Otherwise, there exist two words s, t such that $sut \in x^{-1}Ly^{-1}$. It follows that $xsuty \in L$, a contradiction, since $L \cap A^*uA^* = \emptyset$. Thus $x^{-1}Ly^{-1}$ is nondense. If $L = A^*$, then $x^{-1}Ly^{-1} = A^*$ for all words $x, y \in A^*$. Therefore nondense or full languages are closed under left and right quotients. □

Proposition 4.16. *Nondense or full languages are not closed under inverses of morphisms, even for length-preserving morphisms.*

Proof. Let $\varphi : \{a, b\}^* \to \{a, b\}^*$ be the morphism defined by $\varphi(a) = \varphi(b) = a$. Then a^+ is nondense in $\{a, b\}^*$, but $\varphi^{-1}(a^+) = \{a, b\}^+$ is dense and not full. □

We now give an equational description of nondense or full languages of the form foretold by Theorem 3.10.

Theorem 4.17. *A language of A^* is nondense or full if and only if it satisfies the identity $0 \leqslant x$.*

Proof. Let L be a recognisable language of A^*, let $\eta : A^* \to M$ be its syntactic monoid and P its syntactic image. First observe that the identity $0 \leqslant x$ means that if $0 \in P$, then $P = M$ and hence $M = \{0\}$ since M is the syntactic monoid of P.

Suppose that L is nondense. Then there exists a word $u \in A^*$ such that $L \cap A^*uA^* = \emptyset$. It follows that for all $x \in A^*$, $xu \sim_L u \sim_L ux$ and hence $\eta(u)$ is a zero in M. Furthermore $0 \notin P$. If now L is full, then M is the trivial monoid which has a zero. Thus the identity is satisfied in both cases.

Conversely, assume that M satisfies the identity $0 \leqslant x$. Then M has a zero. If $0 \in P$, then M is the trivial monoid and L is full. Otherwise, $0 \notin P$. Let u be a word such that $\eta(u) = 0$. Then $\eta(A^*uA^*) = 0$ and hence $L \cap A^*uA^* = \emptyset$. Thus L is nondense. □

4.2 Languages defined by density

The *density* of a language $L \subseteq A^*$ is the function which counts the number of words of length n in L. More formally, it is the function $d_L : \mathbb{N} \to \mathbb{N}$ defined by

$$d_L(n) = |L \cap A^n|$$

where $|E|$ denotes the cardinality of a set E. We first state some elementary properties of this function.

Proposition 4.18. *Let L_1 and L_2 be two languages. Then, for all n,*

(1) $d_{L_1 \cup L_2}(n) \leqslant d_{L_1}(n) + d_{L_2}(n)$,
(2) $d_{L_1 \cap L_2}(n) \leqslant \min\{d_{L_1}(n), d_{L_2}(n)\}$,
(3) $d_{L_1 L_2}(n) \leqslant \sum_{0 \leqslant k \leqslant n} d_{L_1}(k) d_{L_2}(n-k)$.

Proof. The double relation

$$(L_1 \cap L_2) \cap A^n \subseteq (L_1 \cup L_2) \cap A^n = (L_1 \cap A^n) \cup (L_2 \cap A^n)$$

gives immediately (1) and (2). Relation (3) follows from the fact that $L_1 L_2 \cap A^n$ is the union of the languages $(L_1 \cap A^k)(L_2 \cap A^{n-k})$, for $0 \leqslant k \leqslant n$. □

If $d_L(n) = O(1)$, then L is called a *slender language*. A language that has at most one word of each length is called a *thin language*. Finally, a language is *sparse* if it has a polynomial density, that is, if $d_L(n) = O(n^k)$ for some $k > 0$. See [177] for a general reference.

Slender languages

Recognisable slender languages have a simple description (see [177, Theorem 3.6]).

Theorem 4.19. *A recognisable language is slender if and only if it is a finite union of languages of the form xu^*y, where $x, u, y \in A^*$.*

Not that if $|A| \leqslant 1$, all recognisable languages are slender. This is trivial if the alphabet is empty. If $A = \{a\}$, every recognisable language of A^* is a finite union of languages of the form xu^*y since $a^k = a^k 1^* 1$, and $a^k a^* = a^k a^* 1$.

Since we only consider lattices of recognisable languages, we are interested in the class of recognisable languages which are either slender or full. We first study their closure properties.

Proposition 4.20. *Recognisable slender languages are closed under finite union, finite intersection, quotients and morphisms.*

Proof. Since any language contained in a slender language is also slender, recognisable slender languages are closed under finite intersection. Closure under

finite union is a direct consequence of Theorem 4.19. Closure under morphisms is trivial.

Let a be a letter of A and let x, u, y be words of A^*. Then

$$a^{-1}(xu^*y) = \begin{cases} (a^{-1}x)u^*y & \text{if } x \neq 1, \\ (a^{-1}u)u^*y & \text{if } x = 1 \text{ and } u \neq 1, \\ a^{-1}y & \text{if } x = 1 \text{ and } u = 1. \end{cases}$$

Therefore $a^{-1}(xu^*y)$ is a recognisable slender language in all cases. It follows now from Theorem 4.19 that recognisable slender languages are closed under left quotient by letters, and, by induction, by any word. The proof for right quotients is similar. □

Corollary 4.21. *Suppose that $|A| \geqslant 2$. Then any recognisable slender language of A^* is nondense.*

Proof. Let L be a recognisable slender language. Then by Proposition 4.20, the language $(A^*)^{-1}L(A^*)^{-1}$ is also slender. Since $|A| \geqslant 2$, this language cannot be equal to A^*. Therefore L is nondense. □

Proposition 4.22. *Recognisable slender languages are not closed under inverses of morphisms, even for length-preserving morphisms.*

Proof. We use the same example as for Proposition 4.16. Let $\varphi : \{a, b\}^* \to \{a, b\}^*$ be the morphism defined by $\varphi(a) = \varphi(b) = a$. Then a^+ is slender in $\{a, b\}^*$, but $\varphi^{-1}(a^+) = \{a, b\}^+$ is nor slender nor full. □

Recall that a *cycle* in a directed graph is a path such that the start vertex and end vertex are the same. A *simple cycle* is a closed directed path, with no repeated vertices other than the starting and ending vertices. The following result is folklore, but we give a self-contained proof for the convenience of the reader.

Theorem 4.23. *Let L be a recognisable language and let $\mathcal{A}$ be its minimal deterministic trimmed automaton. The following conditions are equivalent:*

(1) *L is slender,*
(2) *$\mathcal{A}$ does not contain any connected pair of cycles,*
(3) *$\mathcal{A}$ does not contain any connected pair of simple cycles.*

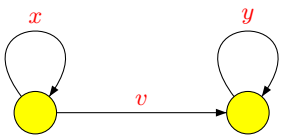


Figure 4.1. Two connected cycles, where $x, y \in A^+$ and $v \in A^*$.

Proof. (1) implies (2). If $\mathcal{A}$ contains a connected pair of cycles, then L contains a language of the form ux^*vy^*w where $u, v, w \in A^*$ and $x, y \in A^+$. In particular, it contains the words $ux^{i|y|}vy^{(n-i)|x|}w$ for $0 \leqslant i \leqslant n$. Therefore $d_L(|uvw| + n|xy|) \geqslant n$ and thus L is not slender.

(2) implies (3) is clear.
(3) implies (1). Let n be the number of states of $\mathcal{A}$ and let w be a word of L of length $\geqslant n$. Consider a successful path p for w. Since $|w| \geqslant n$, this path has a loop, which by (3), is necessarily an iteration of a unique simple cycle, as pictured in Figure 4.2.

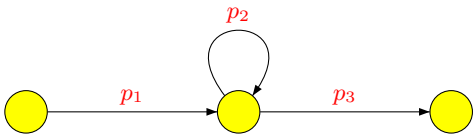


Figure 4.2. The path p.

In other words, one can write $p = p_1 p_2^k p_3$ for some $k > 0$. It follows that L is a subset of $\bigcup_{|u|,|x|,|v|\leqslant n} ux^*v$ and hence is slender. □

We now state the equational characterisation of slender or full languages. We let $i(u)$ denote the first letter (or *initial*) of a word u.

Theorem 4.24. *Suppose that $|A| \geqslant 2$. A recognisable language of A^* is slender or full if and only if it is nondense or full and satisfies the equations $x^\omega u y^\omega = 0$ for each $x, y \in A^+$, $u \in A^*$ such that $i(uy) \neq i(x)$.*

Proof. Let L be a recognisable language of A^*. If L is slender, it is sparse and since $|A| \geqslant 2$, it is nondense by Corollary 4.21. Therefore, by Theorem 4.17, L has a zero and satisfies the equations $0 \leqslant x$ for all $x \in A^*$. It suffices now to prove that if $i(uy) \neq i(x)$, then $x^\omega u y^\omega \leqslant 0$. This formally means that, for every $v, w \in A^*$,

$$\hat{\eta}(vx^\omega u y^\omega w) \in \eta(L) \Rightarrow \eta(v)0\eta(w) \in \eta(L) \tag{4.1}$$

But $\eta(v)0\eta(w) = 0$ and $0 \notin \eta(L)$ since L is a nondense language. It follows that (4.1) holds if and only if $\hat{\eta}(vx^\omega u y^\omega w) \notin \eta(L)$. Let $\mathcal{A} = (Q, A, \cdot, i, F)$ be the minimal trimmed automaton of L. If $\hat{\eta}(vx^\omega u y^\omega w) \in \eta(L)$, then the state $i \cdot vx^\omega u y^\omega w$ is a final state. Setting $p = i \cdot vx^\omega$ and $q = p \cdot uy^\omega$, we have $p \cdot x^\omega = p$ and $q \cdot y^\omega = q$. Furthermore, the condition $i(uy) \neq i(x)$ ensures that the paths defined by x^ω and by uy are distinct. It follows that $\mathcal{A}$ contains a connected pair of cycles, a contradiction to Theorem 4.23.

Suppose now that L is neither slender nor full and let $\mathcal{A}$ be the minimal automaton of L. By Theorem 4.23, $\mathcal{A}$ contains a connected pair of simple cycles. Therefore, there exist some words $x, y \in A^+$, $u_0, u_1, u_2 \in A^*$ such that $i(u_1 y) \neq i(x)$ and $u_0 x^* u_1 y^* u_2 \subseteq L$. It follows that $\hat{\eta}(u_0 x^\omega u_1 y^\omega u_2) \in \eta(L)$ and thus L does not satisfy the equation $x^\omega u_1 y^\omega \leqslant 0$. □

Let us add two comments to this result. First, on a one letter alphabet, every recognisable language is slender, but not necessarily nondense or full: on the alphabet $\{a\}$, the language a^+ is slender, dense, but not full. Second, it looks a little bit suspicious to characterise a class of languages which is not closed under complement by equations of the form $u = v$ and not by equations of the form $u \leqslant v$, as Theorem 3.10 would suggest. The explanation lies in the first condition, "nondense or full", which, by Theorem 4.17, can be defined by equations of the form $0 \leqslant x$.

We now consider the Boolean closure of slender languages. A language is called *coslender* if its complement is slender.

Proposition 4.25. *Recognisable slender or coslender languages are closed under Boolean operations. They are also closed under quotients but not under inverses of morphisms, even for length-preserving morphisms.*

Proof. Let $\mathcal{C}$ be the class of recognisable slender or coslender languages. It is clearly closed under complement. Let $L_1, L_2 \in \mathcal{C}$. If L_1 and L_2 are both slender, then $L_1 \cap L_2$ is also slender by Proposition 4.20. Suppose now that L_1 and L_2 are coslender. Then L_1^c and L_2^c are slender and so is their union by Theorem 4.19. Since $(L_1 \cap L_2)^c = L_1^c \cup L_2^c$, it follows that $L_1 \cap L_2$ is coslender. Thus $\mathcal{C}$ is closed under finite intersection and hence under Boolean operations.

Since slender languages are closed under quotients and since quotients commute with complement (in the sense that $u^{-1}L^c = (u^{-1}L)^c$) $\mathcal{C}$ is closed under quotients.

Finally, let $\varphi : \{a, b, c\}^* \to \{a, b, c\}^*$ be the morphism defined by $\varphi(a) = \varphi(b) = a$ and $\varphi(c) = c$. Then a^+ is slender in $\{a, b\}^*$, but $\varphi^{-1}(a^+) = \{a, b\}^+$ is nor slender nor coslender. □

We now come to the equational characterisation.

Theorem 4.26. *Suppose that $|A| \geqslant 2$. A recognisable language of A^* is slender or coslender if and only if its syntactic monoid has a zero and satisfies the equations $x^\omega u y^\omega = 0$ for each $x, y \in A^+$, $u \in A^*$ such that $i(uy) \neq i(x)$.*

Proof. This is an immediate consequence of Theorem 4.24. □

Note that if $A = \{a\}$, the language $(a^2)^*$ is slender but its syntactic monoid, the cyclic group of order 2, has no zero. Therefore the condition $|A| \geqslant 2$ in Theorem 4.26 is mandatory.

Sparse languages

The closure properties of sparse languages are similar to that of slender languages. See [177, Theorem 3.8].

Proposition 4.27. *Sparse languages are closed under finite union, finite intersection, product and quotients. They are not closed under inverses of morphisms, even for length-preserving morphisms.*

Proof. Proposition 4.18 implies immediately that sparse languages are closed under finite union, finite intersection and product.

Closure under quotients can be proved exactly in the same way as for slender languages and we omit the details. Finally, the example used in the proof of Proposition 4.22 also shows that recognisable slender languages are not closed under inverses of length-preserving morphisms. □

Corollary 4.28. *Suppose that $|A| \geqslant 2$. Then any recognisable sparse language of A^* is nondense.*

Proof. Let L be a recognisable sparse language. Then by Proposition 4.27, the language $(A^*)^{-1}L(A^*)^{-1}$ is also sparse. Therefore this language has a polynomial density and since $|A| \geqslant 2$, it cannot be equal to A^*. Therefore L is nondense. □

Recognisable sparse languages have been characterised in [49, 165, 90]. See also [177, Theorem 3.6]. We gather these results in a slightly different form.

Theorem 4.29. *Let L be a recognisable language. The following conditions are equivalent:*

(1) *L is sparse,*

(2) *L is a finite union of languages of the form $u_0v_1^*u_1\cdots v_k^*u_k$, where u_0, v_1, …, v_k, u_k are words.*

(3) *the minimal deterministic trimmed automaton of L does not contain any pattern of the form*

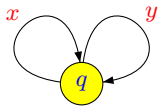


Figure 4.3. The forbidden pattern.

where x and y are nonempty words such that $i(x) \neq i(y)$.

Proof. (2) implies (1). Since recognisable sparse languages are closed under finite union, it suffices to verify that each language L of the form $u_0v_1^*u_1\cdots v_k^*u_k$ is sparse. Considering a typical word $u_0v_1^{r_1}u_1\cdots v_k^{r_k}u_k$ of L, one gets

$$\begin{aligned} d_L(n) &= |\{(r_1,\ldots,r_k) \mid |u_0\cdots u_k| + r_1|v_1| + \ldots + r_k|v_k| = n\})| \\ &\leqslant \mathrm{Card}(\{(n_0,\ldots,n_k) \mid n_0 + n_1 + \ldots + n_k = n\}) \\ &= \binom{n+k}{k} \end{aligned}$$

Since $\binom{n+k}{k} = O(n^k)$, the language L is sparse.

(1) implies (3). Let $\mathcal{A} = (Q, A, \cdot, q_0, F)$ be the minimal deterministic trimmed automaton of L. Suppose that the forbidden pattern occurs in $\mathcal{A}$. Since $\mathcal{A}$ is trimmed, the state q is accessible and co-accessible and there exist two words u and v such that $q_0\cdot u = q$ and $q\cdot v \in F$. It follows that L contains the language $u\{x,y\}^*v$ and a fortiori the language $K = u\{x^{|y|}, y^{|x|}\}^*v$. Setting $r = |x||y|$ and $s = |uv|$, we get $d_K(s + rn) = 2^n$ and thus L is not sparse.

(3) implies (2). If (3) is satisfied, every path of $\mathcal{A}$ contains only elementary loops. It follows that each word of L belongs to some language $u_0v_1^*u_1\cdots v_k^*u_k$ contained in L, where k and the length of the words u_i and v_i are bounded by the number of states of $\mathcal{A}$. There are only finitely many languages of this form and their union is L. Therefore L is sparse. □

It follows that a minimal deterministic automaton recognises a sparse language if and only if it does not contain two cycles reachable from one another.

Corollary 4.30. *Recognisable sparse languages are closed under morphisms.*

Proof. This follows immediately from Condition (2) in the previous theorem. □

Note that the condition *recognisable* in Corollary 4.30 is mandatory. Indeed, consider a bijection f from $\{a,b\}^*$ onto $\mathbb{N}$ such that $f(u) \geqslant |u|$ for all $u \in \{a,b\}^*$. One can construct such a bijection by taking $f(u)$ to be the rank of u in the shortlex order: $1, a, b, aa, ab, ba, bb, aaa, aab, aba, \ldots$. Now, let

$$L = \{uc^{f(|u|)-|u|} \mid u \in \{a,b\}^*\}$$

The property of f implies that L is a thin language. Let now π be the projection morphism from $\{a,b,c\}^*$ onto $\{a,b\}^*$ defined by $\pi(a) = a$, $\pi(b) = b$ and $\pi(c) = 1$. Then $\pi(L) = \{a,b\}^*$ and this language is not sparse.

Theorem 4.31. *Suppose that $|A| \geqslant 2$. A recognisable language of A^* is sparse or full if and only if it is nondense or full and it satisfies the equations $(x^\omega y^\omega)^\omega = 0$ for each $x, y \in A^+$ such that $i(x) \neq i(y)$.*

Proof. The proof is similar to that of Theorem 4.24. Let L be a recognisable language of A^*.

If L is sparse, then it is nondense by Corollary 4.28. Therefore, by Theorem 4.17, L has a zero and satisfies the equations $0 \leqslant x$ for all $x \in A^*$. It suffices now to prove the relation $(x^\omega y^\omega)^\omega \leqslant 0$ when $i(x) \neq i(y)$, which formally means that, for every $v, w \in A^*$,

$$\hat{\eta}(v(x^\omega y^\omega)^\omega w) \in \eta(L) \Rightarrow 0 \in \eta(L) \tag{4.2}$$

But $0 \notin \eta(L)$ since L is a nondense language and thus (4.2) holds if and only if $\hat{\eta}(v(x^\omega y^\omega)^\omega w) \notin \eta(L)$. Let $\mathcal{A} = (Q, A, \cdot, i, F)$ be the minimal trimmed automaton of L. If $\hat{\eta}(v(x^\omega y^\omega)^\omega w) \in \eta(L)$, then the state $i \cdot v(x^\omega y^\omega)^\omega w$ is a final state. Setting $p = i \cdot v(x^\omega y^\omega)^\omega x^\omega$, we get $p \cdot x^\omega = p$ and $p \cdot y^\omega (x^\omega y^\omega)^{\omega-1} = p$ and thus $\mathcal{A}$ contains the pattern of Figure 4.3. This contradicts Theorem 4.29.

If L is not sparse, then its minimal deterministic trimmed automaton contains the pattern represented in Figure 4.3. Consequently, there exist some words $u, v \in A^*$ and $x, y \in A^+$ such that $i(x) \neq i(y)$ and $u\{x,y\}^* v \subseteq L$. It follows that $\hat{\eta}(u(x^\omega y^\omega)^\omega v) \in \eta(L)$ and thus L does not satisfy the equation $u(x^\omega y^\omega)^\omega v \leqslant 0$. □

Pursuing the analogy with slender languages, we consider now the Boolean closure of sparse languages. A language is called *cosparse* if its complement is sparse.

Theorem 4.32. *Suppose that $|A| \geqslant 2$. A recognisable language of A^* is sparse or cosparse if and only if its syntactic monoid has a zero and satisfies the equations $(x^\omega y^\omega)^\omega = 0$ for each $x, y \in A^+$ such that $i(x) \neq i(y)$.*

Proof. This is an immediate consequence of Theorem 4.31. □

5 Exercises

Exercise 1. Let $A = \{a,b\}$ and $L = (a^2)^* b (a^2)^* b$. Compute the syntactic monoid M of L (you should find 12 elements and 6 $\mathcal{J}$-classes). Does L satisfy the the following profinite equations:

(1) $(xy)^\omega = (yx)^\omega$ for all $x, y \in A^*$,
(2) $x^\omega y x^\omega = x^\omega y x^\omega y x^\omega$ for all $x, y \in A^+$,
(3) $x^\omega y^\omega = y^\omega x^\omega$ for all $x, y \in A^*$.

6 Notes

The results presented in Sections 1, 2 and 3 were first presented in [47].

Chapter XIII

Eilenberg's variety theory

1 Streams of languages

A *class of recognisable languages* is a correspondence $\mathcal{F}$ which associates with each alphabet A a set $\mathcal{F}(A^*)$ of recognisable languages of A^* in such a way that, if $\sigma : A \to B$ is a bijection, a language L belongs to $\mathcal{F}(A^*)$ if and only if $\sigma(L)$ belongs to $\mathcal{F}(B^*)$. It follows that, if we fix for each nonnegative integer n an alphabet $A_n = \{a_1, \ldots, a_n\}$, the class $\mathcal{F}$ is completely determined by the family $(\mathcal{F}(A_n^*))_{n\in\mathbb{N}}$.

We use here the terms *class* and *correspondence* instead of *set* and *function* to avoid any paradox of set theory, since it is known, for instance, that the finite sets do not form a set. However, we shall use the term "bijection" instead of "one-to-one and onto correspondence".

A *positive stream of languages* is a class of recognisable languages $\mathcal{V}$ such that

(1) for each alphabet A, $\mathcal{V}(A^*)$ is a lattice of languages,

(2) for each monoid morphism $\varphi : A^* \to B^*$, $X \in \mathcal{V}(B^*)$ implies $\varphi^{-1}(X) \in \mathcal{V}(A^*)$,

A *stream* of languages is a positive stream of languages closed under complementation. This amounts to replacing (1) by (1') in the previous definition

(1′) for each alphabet A, $\mathcal{V}(A^*)$ is a Boolean algebra.

In the sequel, we fix for each nonnegative integer n an alphabet $A_n = \{a_1, \ldots, a_n\}$. Let u and v be two profinite words of $\widehat{A_n^*}$ and let L be a recognisable language of A^*. One says that L *satisfies the profinite identity* $u \to v$ $[u \leftrightarrow v]$ if, for each morphism $\gamma : A_n^* \to A^*$, L satisfies the equation $\widehat{\gamma}(u) \to \widehat{\gamma}(v)$ $[\widehat{\gamma}(u) \leftrightarrow \widehat{\gamma}(v)]$. Recall that, according to Proposition X.2.12, $\widehat{\gamma} : \widehat{A_n^*} \to \widehat{A^*}$ denotes the unique uniformly continuous extension of γ. The next proposition gives an alternative definition.

Proposition 1.1. *Let L be a recognisable language of A^* and let $\varphi : A^* \to M$ be any surjective morphism onto a finite monoid recognising L. Then L satisfies the profinite identity $u \to v$, where $u, v \in \widehat{A_n^*}$, if and only if, for every morphism $\alpha : A_n^* \to M$, $\widehat{\alpha}(u) \in \varphi(L)$ implies $\widehat{\alpha}(v) \in \varphi(L)$.*

Proof. Suppose that, for every morphism $\alpha : A_n^* \to M$, $\widehat{\alpha}(u) \in \varphi(L)$ implies $\widehat{\alpha}(v) \in \varphi(L)$. If $\gamma : A_n^* \to A^*$ is a morphism, then $\widehat{\varphi}(\widehat{\gamma}(u)) \in \varphi(L)$ implies

$\widehat{\varphi}(\widehat{\gamma}(v)) \in \varphi(L)$ and thus L satisfies the equation $\widehat{\gamma}(u) \to \widehat{\gamma}(v)$.

In the opposite direction, suppose that L satisfies the profinite identity $u \to v$ and let $\alpha : A_n^* \to M$ be a morphism. By Corollary II.5.30, there is a morphism $\gamma : A_n^* \to A^*$ such that $\varphi \circ \gamma = \alpha$. Now L satisfies the equation $\widehat{\gamma}(u) \to \widehat{\gamma}(v)$ and thus, by Corollary XII.1.1, $\widehat{\varphi}(\widehat{\gamma}(u)) \in \varphi(L)$ implies $\widehat{\varphi}(\widehat{\gamma}(v)) \in \varphi(L)$. Since $\widehat{\varphi} \circ \widehat{\gamma} = \widehat{\alpha}$, one gets $\widehat{\alpha}(u) \in \varphi(L)$ implies $\widehat{\alpha}(v) \in \varphi(L)$. □

Theorem 1.2. *A class of recognisable languages of A^* is a positive stream of languages if and only if it can be defined by a set of profinite identities of the form $u \to v$. It is a stream of languages if and only if it can be defined by a set of profinite identities of the form $u \leftrightarrow v$.*

Before proving this theorem, let us clarify one point: the identities defining a stream are not necessarily on the same alphabet.

Proof. Let $\mathcal{V}$ be a class of languages defined by a set E of profinite identities of the form $u \to v$. For each n, let E_n be the set of all identities $u \to v$ of E such that $u, v \in \widehat{A}_n^*$. By definition, $\mathcal{V}(A^*)$ is the set of languages satisfying all the equations of the form $\gamma(u) \to \gamma(v)$, where $(u, v) \in E_n$ for some n and $\gamma : A_n^* \to A^*$ is a morphism. In particular, Theorem XII.2.6 shows that $\mathcal{V}(A^*)$ is a lattice of languages.

To prove that $\mathcal{V}$ is a positive stream of languages, it remains to prove that it is closed under inverses of morphisms. Let $\varphi : A^* \to B^*$ be a morphism and let $L \in \mathcal{V}(B^*)$. We claim that $\varphi^{-1}(L)$ satisfies all the identities of E. Indeed, let $(u, v) \in E_n$ and let $\gamma : A_n^* \to A^*$ be a morphism. We want to show that $\varphi^{-1}(L)$ satisfies the equation $\widehat{\gamma(u)} \to \widehat{\gamma(v)}$. Suppose that $\widehat{\gamma}(u) \in \overline{\varphi^{-1}(L)}$. Since Corollary X.3.21 states that $\overline{\varphi^{-1}(L)} = \widehat{\varphi}^{-1}(\overline{L})$, one gets $\widehat{\varphi}(\widehat{\gamma}(u)) \in \overline{L}$. Since $\widehat{\varphi} \circ \widehat{\gamma} = \widehat{\varphi \circ \gamma}$, we finally get $\widehat{\varphi \circ \gamma}(u) \in \overline{L}$. Now, as L satisfies the identity $u \to v$, one obtains $\widehat{\varphi \circ \gamma}(v) \in \overline{L}$, which, by the same argument in reverse order, implies that $\widehat{\gamma}(v) \in \overline{\varphi^{-1}(L)}$. This proves that $\varphi^{-1}(L)$ satisfies the equation $\widehat{\gamma}(u) \to \widehat{\gamma}(v)$ and thereby the identity $u \to v$. This validates the claim and confirms that $\varphi^{-1}(L)$ belongs to $\mathcal{V}(A^*)$. Therefore $\mathcal{V}$ is a positive stream of languages.

Let now $\mathcal{V}$ be a positive stream of languages. Then, for each n, the set $\mathcal{V}(A_n^*)$ is a lattice of languages and by Theorem XII.2.6, it is defined by a set E_n of profinite equations of the form $u \to v$, with $u, v \in \widehat{A}_n^*$. We claim that these equations are actually identities satisfied by $\mathcal{V}$. Let $\gamma : A_n^* \to A^*$ be a morphism and let $u \to v$ be an equation of E_n. If $L \in \mathcal{V}(A^*)$, then $\gamma^{-1}(L) \in \mathcal{V}(A_n^*)$ and thus $\gamma^{-1}(L)$ satisfies the equation $u \to v$. Thus $u \in \overline{\gamma^{-1}(L)}$ implies $v \in \overline{\gamma^{-1}(L)}$. Now, Corollary X.3.21 shows that $\overline{\gamma^{-1}(L)} = \widehat{\gamma}^{-1}(\overline{L})$ and thereby, the conditions $x \in \overline{\gamma^{-1}(L)}$ and $\widehat{\gamma}(x) \in \overline{L}$ are equivalent. Thus $\widehat{\gamma}(u) \in \overline{L}$ implies $\widehat{\gamma}(v) \in \overline{L}$, which means that L satisfies the equation $\widehat{\gamma}(u) \to \widehat{\gamma}(v)$. Therefore $u \to v$ is an identity of $\mathcal{V}$. It follows that $\mathcal{V}$ is defined by a set of identities of the form $u \to v$. □

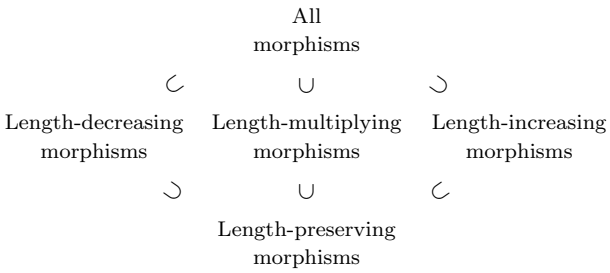
2 $\mathcal{C}$-streams

Let $\mathcal{C}$ be a class of morphisms between finitely generated free monoids that satisfies the following properties :

(1) $\mathcal{C}$ is closed under composition. That is, if A, B and C are finite alphabets, and $f : A^* \to B^*$ and $g : B^* \to C^*$ are elements of $\mathcal{C}$, then $g \circ f$ belongs to $\mathcal{C}$.

(2) $\mathcal{C}$ contains all length-preserving morphisms.

Examples of such classes $\mathcal{C}$ include the classes of all *length-preserving* morphisms (morphisms for which the image of each letter is a letter), of all *length-multiplying* morphisms (morphisms such that, for some integer k, the length of the image of a word is k times the length of the word), all *non-erasing* morphisms (morphisms for which the image of each letter is a nonempty word), all *length-decreasing* morphisms (morphisms for which the image of each letter is either a letter or the empty word) and all morphisms. The following diagram, in which all inclusions are proper, summarizes the relations between these classes.



A *positive $\mathcal{C}$-stream of languages* is a class of recognisable languages $\mathcal{V}$ such that :

(1) for every alphabet A, $\mathcal{V}(A^*)$ is a lattice of languages,

(2) if $\varphi\colon A^* \to B^*$ is a morphism of $\mathcal{C}$, $L \in \mathcal{V}(B^*)$ implies $\varphi^{-1}(L) \in \mathcal{V}(A^*)$,

A *$\mathcal{C}$-stream of languages* is a positive $\mathcal{C}$-stream of languages closed under complement.

Let u and v be two profinite words of $\widehat{A}_n^*$ and let L be a recognisable language of A^*. One says that L *satisfies the profinite $\mathcal{C}$-identity* $u \to v$ $[u \leftrightarrow v]$ if, for each $\mathcal{C}$-morphism $\gamma : A_n^* \to A^*$, L satisfies the equation $\hat{\gamma}(u) \to \hat{\gamma}(v)$ $[\hat{\gamma}(u) \leftrightarrow \hat{\gamma}(v)]$.

Theorem 2.3. *A class of recognisable languages of A^* is a positive $\mathcal{C}$-stream of languages if and only if it can be defined by a set of profinite $\mathcal{C}$-identities of the form $u \to v$. It is a stream of languages if and only if it can be defined by a set of profinite $\mathcal{C}$-identities of the form $u \leftrightarrow v$.*

Proof. The proof of Theorem 1.2 carries over by changing every occurrence of "morphism" by "$\mathcal{C}$-morphism". Note however that the closure of $\mathcal{C}$ under composition is needed to show that $\gamma \circ \varphi$ belongs to $\mathcal{C}$. □

The notion of $\mathcal{C}$-identity can be sometimes confusing. Let us illustrate it by a few examples.

Example 2.1. A recognisable language L satisfies the identity [length-preserving, non-erasing, length-multiplying identity] $xyx \leqslant yxy$ if, for all words [letters, nonempty words, words of the same length] x and y, the relation $xyx \leqslant_L yxy$ holds.

3 Varieties of languages

A *positive variety of languages* is a class of recognisable languages such that

(1) for each alphabet A, $\mathcal{V}(A^*)$ is a lattice of languages,

(2) for each monoid morphism $\varphi : A^* \to B^*$, $X \in \mathcal{V}(B^*)$ implies $\varphi^{-1}(X) \in \mathcal{V}(A^*)$,

(3) If $X \in \mathcal{V}(A^*)$ and $u \in A^*$, $u^{-1}X \in \mathcal{V}(A^*)$ and $Xu^{-1} \in \mathcal{V}(A^*)$.

A *variety* of languages is a positive variety of languages closed under complementation. This amounts to replacing (1) by (1′) in the previous definition

(1′) for each alphabet A, $\mathcal{V}(A^*)$ is a Boolean algebra.

Let u and v be two profinite words of $\widehat{A}_n^*$ and let L be a recognisable language of A^*. One says that L *satisfies the profinite identity* $u \leqslant v$ $[u = v]$ if, for all morphisms $\gamma : A_n^* \to A^*$, L satisfies the equation $\hat{\gamma}(u) \leqslant \hat{\gamma}(v)$ $[\hat{\gamma}(u) = \hat{\gamma}(v)]$.

Theorem 3.4. *A class of recognisable languages of A^* is a positive variety of languages if and only if it can be defined by a set of profinite identities of the form $u \leqslant v$. It is a variety of languages if and only if it can be defined by a set of profinite identities of the form $u = v$.*

Proof. The proof, which combines the arguments of Theorems XII.3.10 and 1.2 is left as an exercise to the reader. □

A *positive $\mathcal{C}$-variety of languages* is a class of recognisable languages $\mathcal{V}$ such that

(1) for every alphabet A, $\mathcal{V}(A^*)$ is a lattice of languages,

(2) if $\varphi\colon A^* \to B^*$ is a morphism of $\mathcal{C}$, $L \in \mathcal{V}(B^*)$ implies $\varphi^{-1}(L) \in \mathcal{V}(A^*)$,

(3) if $L \in \mathcal{V}(A^*)$ and if $a \in A$, then $a^{-1}L$ and La^{-1} are in $\mathcal{V}(A^*)$.

A *$\mathcal{C}$-variety of languages* is a positive $\mathcal{C}$-variety of languages closed under complement.

Let u and v be two profinite words of $\widehat{A}_n^*$ and let L be a recognisable language of A^*. One says that L *satisfies the profinite $\mathcal{C}$-identity* $u \leqslant v$ $[u = v]$ if, for all $\mathcal{C}$-morphisms $\gamma : A_n^* \to A^*$, L satisfies the equation $\hat{\gamma}(u) \leqslant \hat{\gamma}(v)$ $[\hat{\gamma}(u) = \hat{\gamma}(v)]$.

Theorem 3.5. *A class of recognisable languages of A^* is a positive $\mathcal{C}$-variety of languages if and only if it can be defined by a set of profinite identities of the form $u \leqslant v$. It is a $\mathcal{C}$-variety of languages if and only if it can be defined by a set of profinite identities of the form $u = v$.*

Proof. The proof, which combines the arguments of Theorems XII.3.10 and 2.3 is left as an exercise to the reader. □

When $\mathcal{C}$ is the class of all morphisms, we recover the definition of a variety of languages. When $\mathcal{C}$ is the class of length-preserving [length-multiplying, nonerasing, length-decreasing] morphisms, we use the term *lp*-variety [*lm*-variety, *ne*-variety, *ld*-variety] of languages.

4 The variety theorem

In this section, we present a slightly different algebraic point of view to characterise varieties of languages, proposed by Eilenberg [42].

If $\mathbf{V}$ is a variety of finite monoids, let $\mathcal{V}(A^*)$ denote the set of recognisable languages of A^* whose syntactic monoid belongs to $\mathbf{V}$. The following is an equivalent definition:

Proposition 4.6. *$\mathcal{V}(A^*)$ is the set of languages of A^* recognised by a monoid of $\mathbf{V}$.*

Proof. If $L \in \mathcal{V}(A^*)$, then the syntactic monoid of L, which recognises L, belongs to $\mathbf{V}$. Conversely, if L is recognised by a monoid M of $\mathbf{V}$, then by Proposition IV.4.26, the syntactic monoid of L divides M and thus belongs also to $\mathbf{V}$. □

The correspondence $\mathbf{V} \to \mathcal{V}$ associates with each variety of finite monoids a class of recognisable languages. We shall see later (Proposition 4.9) that $\mathcal{V}$ is a variety of languages. For now, we show that this correspondence is one-to-one.

Theorem 4.7. *Let $\mathbf{V}$ and $\mathbf{W}$ be two varieties of finite monoids. Suppose that $\mathbf{V} \to \mathcal{V}$ and $\mathbf{W} \to \mathcal{W}$. Then $\mathbf{V} \subseteq \mathbf{W}$ if and only if, for every finite alphabet A, $\mathcal{V}(A^*) \subseteq \mathcal{W}(A^*)$. In particular, $\mathbf{V} = \mathbf{W}$ if and only if $\mathcal{V} = \mathcal{W}$.*

Proof. If $\mathbf{V} \subseteq \mathbf{W}$, it follows immediately from the definitions that $\mathcal{V}(A^*) \subseteq \mathcal{W}(A^*)$. The proof of the opposite direction of the theorem is based on the following proposition.

Proposition 4.8. *Let $\mathbf{V}$ be a variety of monoids and let $M \in \mathbf{V}$. Then there exist a finite alphabet A and languages $L_1, \ldots, L_k \in \mathcal{V}(A^*)$ such that M is isomorphic to a submonoid of $M(L_1) \times \cdots \times M(L_k)$.*

Proof. Since M is finite, there exists a finite alphabet A and a surjective monoid morphism $\varphi : A^* \to M$. For each $s \in M$, put $L_s = \varphi^{-1}(s)$. Then L_s is recognised by M and thus $L_s \in \mathcal{V}(A^*)$. Let M_s be the syntactic monoid of L_s. By Proposition IV.4.25, it is also the syntactic monoid of the singleton $\{s\}$ in M. We let $\pi_s : M \to M_s$ denote the projection and by $\pi : M \to \prod_{s \in M} M_s$ the morphism of monoids defined by $\pi(x) = (\pi_s(x))_{s \in M}$. We claim that π is injective. If $\pi(x) = \pi(y)$, then in particular, $\pi_y(x) = \pi_y(y)$. This means that, for every $s, t \in M$, $syt = y$ if and only if $sxt = y$. Applying this result with $s = t = 1$, one gets $x = y$. This proves the claim and shows that M is isomorphic to a submonoid of $\prod_{s \in M} M_s$. □

We can now complete the proof of Theorem 4.7. Suppose that $\mathcal{V}(A^*) \subseteq \mathcal{W}(A^*)$ for every finite alphabet A and let $M \in \mathbf{V}$. Then by Proposition 4.8, M is isomorphic to a submonoid of the form $M(L_1) \times \cdots \times M(L_k)$, where $L_1, \ldots, L_k \in \mathcal{V}(A^*)$. It follows that $L_1, \ldots, L_k \in \mathcal{W}(A^*)$ and hence $M(L_1), \ldots, M(L_k) \in \mathbf{W}$. Therefore $M \in \mathbf{W}$. □

We now characterise the classes of languages which can be associated with a variety of monoids.

Proposition 4.9. *Let $\mathbf{V}$ be a variety of finite monoids. If $\mathbf{V} \to \mathcal{V}$, then $\mathcal{V}$ is a variety of languages.*

Proof. Let $L, L_1, L_2 \in \mathcal{V}(A^*)$ and let $a \in A$. Then by definition $M(L)$, $M(L_1)$, $M(L_2)$ are in $\mathbf{V}$. By Proposition IV.2.10, the languages $L_1 \cup L_2$ and $L_1 \cap L_2$ are recognised by a submonoid T of $M(L_1) \times M(L_2)$. Now since $\mathbf{V}$ is a variety of monoids, $T \in \mathbf{V}$ and thus $L_1 \cup L_2$ and $L_1 \cap L_2$ belong to $\mathcal{V}(A^*)$. Since $\emptyset$

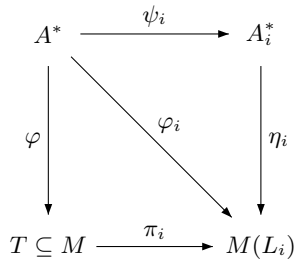
and A^* are recognised by the trivial monoid, which is certainly in $\mathbf{V}$, $\mathcal{V}(A^*)$ is a Boolean algebra of languages. Similarly, Proposition IV.2.12 shows that the languages $a^{-1}L$ and La^{-1} are recognised by $M(L)$ and Proposition IV.2.11 shows that, if $\varphi : B^* \to A^*$ is a monoid morphism, then $\varphi^{-1}(L)$ is recognised by $M(L)$. Thus $\mathcal{V}$ is a variety of languages. □

To each variety of languages $\mathcal{V}$, we associate the variety of monoids $\mathbf{V}$ generated by the monoids of the form $M(L)$ where $L \in \mathcal{V}(A^*)$ for a certain alphabet A. This defines a correspondence $\mathcal{V} \to \mathbf{V}$. We are now ready to state Eilenberg's variety theorem.

Theorem 4.10. *The correspondences $\mathbf{V} \to \mathcal{V}$ and $\mathcal{V} \to \mathbf{V}$ define mutually inverse bijective correspondences between varieties of finite monoids and varieties of languages.*

Proof. Let $\mathcal{V}$ be a variety of languages and suppose that $\mathcal{V} \to \mathbf{V}$ and $\mathbf{V} \to \mathcal{W}$. We claim that $\mathcal{V} = \mathcal{W}$. First, if $L \in \mathcal{V}(A^*)$, one has $M(L) \in \mathbf{V}$ by definition and therefore $L \in \mathcal{W}(A^*)$, still by definition. Therefore, for every alphabet A, $\mathcal{V}(A^*) \subseteq \mathcal{W}(A^*)$.

The inclusion $\mathcal{W}(A^*) \subseteq \mathcal{V}(A^*)$ is more difficult to prove. Let $L \in \mathcal{W}(A^*)$. Then $M(L) \in \mathbf{V}$ and since $\mathbf{V}$ is the variety generated by the monoids of the form $M(L)$ where L is a language of $\mathcal{V}$, there exist an integer $n > 0$ and, for $1 \leqslant i \leqslant n$, alphabets A_i and languages $L_i \in \mathcal{V}_i(A_i^*)$ such that $M(L)$ divides $M(L_1) \times \cdots \times M(L_n)$. Denote this product of monoids by M. Since $M(L)$ divides M, it is a quotient of a submonoid T of M. By Corollary IV.4.23, T recognises L. Therefore there exists a surjective morphism of monoids $\varphi : A^* \to T$ and a subset P of T such that $L = \varphi^{-1}(P)$. Let $\pi_i : M \to M(L_i)$ be the i-th projection defined by $\pi_i(s_1, \ldots, s_n) = s_i$. Put $\varphi_i = \pi_i \circ \varphi$ and let $\eta_i : A_i^* \to M(L_i)$ be the syntactic morphism of L_i. Since η_i is onto, there exists by Corollary II.5.30 a morphism of monoids $\psi_i : A^* \to A_i^*$ such that $\varphi_i = \eta_i \circ \psi_i$. We can summarise the situation in a diagram:



We recall that we are seeking to prove that $L \in \mathcal{V}(A^*)$, which is finally obtained by a succession of reductions of the problem. First, one has

$$L = \varphi^{-1}(P) = \bigcup_{s \in P} \varphi^{-1}(s)$$

Since $\mathcal{V}(A^*)$ is closed under union, it suffices to establish that for every $s \in P$, one has $\varphi^{-1}(s)$ belongs to $\mathcal{V}(A^*)$. Setting $s = (s_1, \ldots, s_n)$, one gets $\{s\} = \bigcap_{1 \leqslant i \leqslant n} \pi_i^{-1}(s_i)$. Consequently

$$\varphi^{-1}(s) = \bigcap_{1 \leqslant i \leqslant n} \varphi^{-1}(\pi_i^{-1}(s_i)) = \bigcap_{1 \leqslant i \leqslant n} \varphi_i^{-1}(s_i)$$

As $\mathcal{V}(A^*)$ is a lattice, it suffices to establish that, for $1 \leqslant i \leqslant n$, $\varphi_i^{-1}(s_i) \in \mathcal{V}(A^*)$.

Since $\varphi_i = \eta_i \circ \psi_i$, one has $\varphi_i^{-1}(s_i) = \psi_i^{-1}(\eta_i^{-1}(s_i))$. Now since $\mathcal{V}$ is a variety of languages, it suffices to prove that $\eta_i^{-1}(s_i) \in \mathcal{V}(A_i^*)$, which results from the following lemma.

Lemma 4.11. *Let $\mathcal{V}$ be a variety of languages and let $\eta : A^* \to M$ be the syntactic morphism of a language L of $\mathcal{V}(A^*)$. Then for every $x \in M$, $\eta^{-1}(x)$ belongs to $\mathcal{V}(A^*)$.*

Proof. Let $P = \eta(L)$. Then $L = \eta^{-1}(P)$. Setting $E = \{(s,t) \in M^2 \mid sxt \in P\}$, we claim that

$$\{x\} = \Big(\bigcap_{(s,t)\in E} s^{-1}Pt^{-1}\Big) - \Big(\bigcup_{(s,t)\in E^c} s^{-1}Pt^{-1}\Big) \tag{4.1}$$

Let R be the right-hand side of (4.1). It is clear that x belongs to R. Conversely, let u be an element of R. If $sxt \in P$, then $u \in s^{-1}Pt^{-1}$, that is, $sut \in P$. And if $sxt \notin P$, then $u \notin s^{-1}Pt^{-1}$. It follows that $u \sim_L x$ and thus $u = x$, which proves the claim. Since η^{-1} commutes with Boolean operations and quotients, $\eta^{-1}(x)$ is a Boolean combination of quotients of L and hence belongs to $\mathcal{V}(A^*)$. □

There is an analoguous theorem for varieties of ordered monoids.

Theorem 4.12. *The correspondences $\mathbf{V} \to \mathcal{V}$ and $\mathcal{V} \to \mathbf{V}$ define mutually inverse bijective correspondences between varieties of finite ordered monoids and positive varieties of languages.*

Eilenberg's +-varieties

In his original presentation, Eilenberg introduced a variant of his variety theorem. We shall briefly review this version and discuss its connection with the point of view presented in this chapter.

The first difference is that languages are now subsets of a free semigroup, instead of a free monoid in the usual case. In this setting, it is natural to use semigroups instead of monoids and in particular, there is a notion of syntactic semigroup. A *+-variety of languages* is a class of recognisable languages such that

(1) for each alphabet A, $\mathcal{V}(A^+)$ is a Boolean algebra,
(2) for each semigroup morphism $\varphi : A^+ \to B^+$, if $X \in \mathcal{V}(B^+)$ then $\varphi^{-1}(X) \in \mathcal{V}(A^+)$,
(3) If $X \in \mathcal{V}(A^+)$ and $u \in A^*$, $u^{-1}X \in \mathcal{V}(A^+)$ and $Xu^{-1} \in \mathcal{V}(A^+)$.

Positive +-varieties of languages can be defined in the same way. The second Eilenberg's

Theorem 4.13. *The correspondences $\mathbf{V} \to \mathcal{V}$ and $\mathcal{V} \to \mathbf{V}$ define mutually inverse bijective correspondences between varieties of finite semigroups and +-varieties of languages.*

There is an analoguous theorem for varieties of ordered semigroups.

Theorem 4.14. *The correspondences $\mathbf{V} \to \mathcal{V}$ and $\mathcal{V} \to \mathbf{V}$ define mutually inverse bijective correspondences between varieties of finite ordered semigroups and positive +-varieties of languages.*

To interpret this result in the context of this chapter, let us introduce a notation: if S is a semigroup, let S^I denote the monoid $S \cup \{I\}$, where I is a new identity.

Let $\mathbf{V}$ be a variety of finite semigroups and let $\mathcal{V}$ be the corresponding variety of languages. We define another class of languages $\mathcal{V}'$ as follows. For each alphabet A, $\mathcal{V}'(A^*)$ consists of all languages of A^* recognised by a morphism $\varphi : A^* \to S^I$ such that the semigroup $\varphi(A^+)$ is in $\mathbf{V}$. This is the case in particular if $S \in \mathbf{V}$ and $\varphi : A^+ \to S$ is a surjective morphism. Then one can show that $\mathcal{V}$ is a *ne*-variety of languages. Moreover, every language of $\mathcal{V}(A^+)$ is a language of $\mathcal{V}'(A^*)$. Conversely, if L is a language of $\mathcal{V}'(A^*)$, then $L \cap A^+$ is a language of $\mathcal{V}(A^+)$.

5 Summary

We summarise in a table the various types of equations we have used so far.

Closed under	**Equations**	**Definition**
$\cup, \cap$	$u \to v$	$\widehat{\eta}(u) \in \widehat{\eta}(L) \Rightarrow \widehat{\eta}(v) \in \widehat{\eta}(L)$
quotient	$u \leqslant v$	$xuy \to xvy$
complement	$u \leftrightarrow v$	$u \to v$ and $v \to u$
quotient and complement	$u = v$	$xuy \leftrightarrow xvy$
Closed under inverses of morphisms		**Interpretation of variables**
all morphisms		words
nonerasing morphisms		nonempty words
length multiplying morphisms		words of equal length
length preserving morphisms		letters

6 Notes

Varieties of languages and the variety theorem (Section 4) are due to Eilenberg [42]. C-varieties were introduced by Straubing [162] and under a slightly different form, by Ésik [44].

Chapter XIV

Algebraic characterisations

In this chapter, we give explicit examples of the algebraic characterisations described in Chapter X.

1 Varieties of languages

The two simplest examples of varieties of languages are the variety $\mathcal{I}$ corresponding to the trivial variety of monoids $\mathbf{1}$ and the variety $\mathcal{R}at$ of rational languages corresponding to the variety of all monoids. For each alphabet A, $\mathcal{I}(A^*)$ consists of the empty and the full language A^* and $\mathcal{R}at(A^*)$ is the set of all rational languages on the alphabet A.

We provide the reader with many more examples in this section. Other important examples are the topic of Chapters VI, VII and XVII.

1.1 Locally finite varieties of languages

Let $\mathcal{V}$ be a [positive] variety of languages. One says that $\mathcal{V}$ is *locally finite* if the corresponding variety of [ordered] monoids is locally finite. Here is a more combinatorial characterisation.

Proposition 1.1. *A [positive] variety of languages $\mathcal{V}$ is locally finite if and only if, for each alphabet A, $\mathcal{V}(A^*)$ is a finite set.*

Proof. Let $\mathbf{V}$ be the variety of [ordered] monoids corresponding to $\mathcal{V}$. If $\mathbf{V}$ is locally finite, then, for each alphabet A, $\widehat{F}_{\mathbf{V}}(A)$ is a finite monoid which recognises all the languages of $\mathcal{V}(A^*)$. It follows that $\mathcal{V}(A^*)$ is a finite set.

Conversely, suppose that for each alphabet A, $\mathcal{V}(A^*)$ is a finite set. Define an equivalence $\equiv_{\mathbf{V}}$ on A^* by $u \equiv_{\mathbf{V}} v$ if and only if, for all $L \in \mathcal{V}(A^*)$, the conditions $u \in L$ and $v \in L$ are equivalent. Since $\mathcal{V}(A^*)$ is finite, this equivalence has finite index. We claim that $u \equiv_{\mathbf{V}} v$ implies $u \sim_{\mathbf{V}} v$. Indeed, let φ be a morphism from A^* onto a monoid M of $\mathbf{V}$ and suppose that $u \equiv_{\mathbf{V}} v$. Then the language $L = \varphi^{-1}(\varphi(u))$ belongs to $\mathcal{V}(A^*)$ and since $u \in L$, one gets $v \in L$, that is $\varphi(v) = \varphi(u)$, which proves the claim. It follows that $\sim_{\mathbf{V}}$ has finite index and thus $\widehat{F}_{\mathbf{V}}(A)$ is a finite monoid. □

A combinatorial description

Proposition XI.2.8 shows that if $\mathbf{V}$ is generated by a single [ordered] monoid, then $\mathcal{V}$ is locally finite. In this case, one can give a more precise description of $\mathcal{V}$.

Proposition 1.2. *Let $\mathbf{V}$ be a variety of ordered monoids generated by a single ordered monoid M and let $\mathcal{V}$ be the corresponding positive variety. Then, for each alphabet A, $\mathcal{V}(A^*)$ is the lattice generated by the sets of the form $\varphi^{-1}(\uparrow m)$, where $\varphi : A^* \to M$ is an arbitrary morphism and $m \in M$.*

Proof. It is clear that $\varphi^{-1}(\uparrow m) \in \mathcal{V}(A^*)$ and thus $\mathcal{V}(A^*)$ also contains the lattice generated by these sets. Conversely, let $L \in \mathcal{V}(A^*)$. Then there exists an integer $n \geqslant 0$, a morphism $\varphi : A^* \to M^n$ and an upper set I of M^n such that $L = \varphi^{-1}(I)$. Since $\varphi^{-1}(I) = \bigcup_{m \in P} \varphi^{-1}(\uparrow m)$, it is sufficient to establish the result when $L = \varphi^{-1}(\uparrow m)$ where $m \in M^n$. Denote by π_i the i-th projection from M^n onto M. Setting $m = (m_1, \ldots, m_n)$, we have $m = \bigcap_{1 \leqslant i \leqslant n} \pi_i^{-1}(m_i)$, whence

$$\varphi^{-1}(\uparrow m) = \bigcap_{1 \leqslant i \leqslant n} (\pi_i \circ \varphi)^{-1}(\uparrow m_i)$$

Since $m_i \in M$ and $\pi_i \circ \varphi$ is a morphism from A^* to M, the result follows. □

There is of course a similar result for the varieties of monoids, the proof of which is similar.

Proposition 1.3. *Let $\mathbf{V}$ be a variety of monoids generated by a single monoid M and let $\mathcal{V}$ be the corresponding variety of languages. Then, for every alphabet A, $\mathcal{V}(A^*)$ is the Boolean algebra generated by the sets of the form $\varphi^{-1}(m)$, where $\varphi : A^* \to M$ is an arbitrary morphism and $m \in M$.*

Languages corresponding to $\mathbf{J_1}$, $\mathbf{J_1^+}$ and $\mathbf{J_1^-}$

Let $\mathcal{J}_1$ [$\mathcal{J}_1^+$, $\mathcal{J}_1^-$] denote the [positive] variety of languages corresponding to $\mathbf{J_1}$ [$\mathbf{J_1^+}$, $\mathbf{J_1^-}$].

Proposition 1.4. *For each alphabet A, $\mathcal{J}_1(A^*)$ is the Boolean algebra generated by the languages of the form A^*aA^* where a is a letter. Equivalently, $\mathcal{J}_1(A^*)$ is the Boolean algebra generated by the languages of the form B^* where B is a subset of A.*

Proof. The equality of the two Boolean algebras considered in the statement results from the formulas

$$B^* = A^* - \bigcup_{a \in A - B} A^*aA^* \quad \text{and} \quad A^*aA^* = A^* - (A - \{a\})^*$$

Since $\mathbf{J_1}$ is generated by U_1, one can use Proposition 1.3 to describe $\mathcal{J}_1$. Let $\varphi : A^* \to U_1$ be a morphism, and let $B = \{a \in A \mid \varphi(a) = 1\}$. Then $\varphi^{-1}(1) = B^*$ and $\varphi^{-1}(0) = A^* - B^*$, which establishes the proposition. □

If B is a subset of A, let $F(B)$ denote the set of words of A^* containing at least one occurrence of each letter of B. Thus

$$F(B) = \bigcap_{a \in B} A^* a A^*$$

The next proposition is thus a variant of Proposition 1.4 and its proof is left as an exercise to the reader.

Proposition 1.5. *For each alphabet A, $\mathcal{J}_1^+(A^*)$ is the set of finite unions of languages of the form $F(B)$ where $B \subseteq A$. Similarly, $\mathcal{J}_1^-(A^*)$ is the set of finite unions of languages of the form B^* where $B \subseteq A$.*

Languages corresponding to $\mathbf{R_1}$ and $\mathbf{L_1}$

Another interesting example of a locally finite variety is the variety $\mathbf{R_1}$ of idempotent and $\mathcal{R}$-trivial monoids.

Proposition 1.6. *Let L be a recognisable subset of A^* and let M be its syntactic monoid. The following conditions are equivalent:*

(1) *M divides $\tilde{U}_2^n$ for some $n > 0$,*
(2) *M belongs to $\mathbf{R_1}$,*
(3) *M satisfies the identity $xyx = xy$,*
(4) *L is a disjoint union of sets of the form*

$$a_1\{a_1\}^* a_2\{a_1, a_2\}^* a_3\{a_1, a_2, a_3\}^* \cdots a_n\{a_1, a_2, \ldots, a_n\}^*$$

where the a_i's are distinct letters of A,

(5) *L is a Boolean combination of sets of the form B^*aA^*, where $a \in A$ and $B \subseteq A$.*

Proof. (1) implies (2) since $\tilde{U}_2 \in \mathbf{R_1}$ and $\mathbf{R_1}$ is a variety.
(2) implies (3). Let $x, y \in M$. Since M is idempotent, $xy = xyxy$ and thus $xy \mathcal{R} xyx$. But M is $\mathcal{R}$-trivial and therefore $xy = xyx$.
(3) implies (4). Let $\rho : A^* \to A^*$ be the function which associates with any word u the sequence of all distinct letters of u in the order in which they first appear when u is read from left to right. For example, if $u = caabacb$, then $\rho(u) = cab$. In fact ρ is a sequential function, realised by the sequential transducer $\mathcal{T} = (\mathcal{P}(A), A, \emptyset, \cdot, *)$, where the transition and the output functions are defined by

$$B \cdot a = B \cup \{a\}$$
$$B * a = \begin{cases} 1 & \text{if } a \in B \\ 0 & \text{otherwise} \end{cases}$$

Define an equivalence $\sim$ on A^* by setting $u \sim v$ if $\rho(u) = \rho(v)$. It is easy to see that the equivalence classes of $\sim$ are the disjoint sets

$$L_{(a_1,\ldots,a_n)} = a_1\{a_1\}^* a_2\{a_1, a_2\}^* a_3\{a_1, a_2, a_3\}^* \cdots a_n\{a_1, a_2, \ldots, a_n\}^*$$

where $(a_1, \ldots, a_n)$ is a sequence of distinct letters of A. We claim that $\sim$ is a congruence. If $u \sim v$, then u and v belong to some set $L_{(a_1,\ldots,a_n)}$. Let a be a letter. If $a = a_i$ for some i, then $ua, va \in L_{(a_1,\ldots,a_n)}$, and $au, av \in$

$L_{(a,a_1,\ldots,a_{i-1},a_{i+1},\ldots,a_n)}$. Thus $ua \sim va$ and $au \sim av$. If $a \notin \{a_1,\ldots,a_n\}$, then $ua, va \in L_{(a_1,\ldots,a_n,a)}$ and $au, av \in L_{(a,a_1,\ldots,a_n)}$ and thus again $ua \sim va$ and $au \sim av$, which proves the claim.

Let $\eta : A^* \to M$ be the syntactic morphism of L. If $u \in L_{(a_1,\ldots,a_n)}$, then $u = a_1u_1a_2u_2\cdots a_nu_n$ where $u_i \in \{a_1,\ldots,a_i\}^*$ for $1 \leqslant i \leqslant n$ and thus by (3), $\eta(u) = \eta(a_1\cdots a_n)$. It follows that $u \sim v$ implies $\eta(u) = \eta(v)$ and therefore L is a disjoint union of equivalence classes of $\sim$, that is of sets of the form $L_{(a_1,\ldots,a_n)}$. (4) implies (5). First observe that

$$L_{(a_1,\ldots,a_n)} = A_n^* \cap \bigcap_{1\leqslant i\leqslant n} A_{i-1}^* a_i A^* \text{ where } A_i = \{a_1,\ldots,a_i\} \text{ and } A_0 = \emptyset$$

Condition (5) is now a consequence of the following equalities:

$$A_i^* = A^* - \bigcup_{a\notin A_i} A^*aA^* \qquad A_{i-1}^* a_i A_i^* = A_{i-1}^* a_i A^* \cap A_i^*$$

(5) implies (1). By the variety theorem (Theorem XIII.4.7), it is sufficient to show that, for $B \subseteq A$ and $a \in A$, B^*aA^* is recognised by $\tilde{U}_2$. Let $\tilde{U}_2 = \{1, a_1, a_2\}$ and let $\varphi : A^* \to \tilde{U}_2$ be the morphism defined by

$$\varphi(a) = a_1$$
$$\varphi(b) = \begin{cases} 1 & \text{if } b \in B - \{a\} \\ a_2 & \text{for } b \in A - (B \cup \{a\}) \end{cases}$$

Then $\varphi^{-1}(a_1) = B^*aA^*$, which concludes the proof. □

There is of course a dual version for the variety $\mathbf{L_1}$ of idempotent and $\mathcal{L}$-trivial monoids.

Proposition 1.7. *Let L be a recognisable subset of A^* and let M be its syntactic monoid. The following conditions are equivalent:*

(1) *M divides U_2^n for some $n > 0$,*
(2) *M belongs to $\mathbf{L_1}$,*
(3) *M satisfies the identity $xyx = yx$,*
(4) *L is a disjoint union of sets of the form*

$$\{a_1, a_2, \ldots, a_n\}^* a_n \{a_1, a_2, \ldots, a_{n-1}\} \cdots \{a_1, a_2\}^* a_2 \{a_1\}^* a_1$$

where the a_i's are distinct letters of A,

(5) *L is a Boolean combination of sets of the form A^*aB^*, where $a \in A$ and $B \subseteq A$.*

1.2 Commutative languages

We now come to the study of commutative languages. We study successively the languages corresponding to aperiodic and commutative monoids (variety **Acom**), to commutative groups and to arbitrary commutative monoids.

We denote by $[u]$ the *commutative closure* of a word u, which is the set of words commutatively equivalent to u. For instance, $[aab] = \{aab, aba, baa\}$. A language L is *commutative* if, for every word $u \in L$, $[u]$ is contained in L. Equivalently, a language is *commutative* if its syntactic monoid is commutative.

Languages recognised by aperiodic commutative monoids

If a is a letter of an alphabet A, let $L(a,k)$ denote the set of words of A^* which contain exactly k occurrences of a

$$L(a,k) = \{u \in A^+ \mid |u|_a = k\}$$

Then the following result holds.

Proposition 1.8. *For each alphabet A, $\mathcal{A}com(A^*)$ is the Boolean algebra generated by the sets of the form $L(a,k)$ where $a \in A$ and $k \geqslant 0$.*

Proof. First, every set of the form $L(a,k)$ is recognised by an aperiodic commutative monoid. Indeed, let $N = \{1, x, x^2, \ldots, x^k, x^{k+1}\}$ be the cyclic monoid defined by the relation $x^{k+2} = x^{k+1}$, and let $\varphi : A^* \to N$ be the morphism defined by $\varphi(a) = x$ and $\varphi(b) = 1$ if $b \neq a$. Then clearly $L(a,k) = \varphi^{-1}(x^k)$.

By Proposition XI.4.21, **Acom** is generated by its cyclic monoids, and Proposition 1.3 can be used to describe $\mathcal{A}com$. Let $M = \{1, x, x^2, \ldots, x^n\}$ be a cyclic monoid, defined by the relation $x^{n+1} = x^n$, and let $\varphi : A^* \to M$ be a morphism. Then for each $a \in A$ there exists an integer n_a such that $\varphi(a) = x^{n_a}$. Let k be an integer such that $0 \leqslant k < n$. Then

$$\begin{aligned}\varphi^{-1}(x^k) &= \{u \in A^* \mid \sum_{a \in A} n_a |u|_a = k\} \\ &= \bigcup \bigcap_{a \in A} L(a, k_a)\end{aligned}$$

where the union is taken over the set of families $(k_a)_{a \in A}$ such that $\sum_{a \in A} n_a k_a = k$. Finally, for $k = n$, we have

$$\varphi^{-1}(x^n) = A^* - \bigcup_{0 \leqslant k < n} \varphi^{-1}(x^k)$$

which concludes the proof. □

Languages recognised by commutative groups

For each positive integer n, let $\mathbf{Ab}(n)$ be the variety of all abelian groups of exponent dividing n. This variety is known to be generated by the cyclic groups of order n. Let us call *n-commutative* a language recognised by a group in $\mathbf{Ab}(n)$. A description of these languages was given in [42].

Proposition 1.9. *For each alphabet A, the n-commutative languages of A^* form the Boolean algebra generated by the languages of the form*

$$F(a,k,n) = \{u \in A^* \mid |u|_a \equiv k \bmod n\} = ((B^*a)^n)^*(B^*a)^k B^*,$$

where $a \in A$, $B = A - \{a\}$ and $0 \leqslant k < n$.

We shall present an improved version of this result, which avoids using complementation. Let $A = \{a_1, \ldots, a_s\}$ be an alphabet. Let us call *elementary n-commutative* a language of the form

$$F(r_1, \ldots, r_s, n) = \{u \in A^* \mid |u|_{a_1} \equiv r_1 \bmod n, \ \ldots, \ |u|_{a_s} \equiv r_s \bmod n\}$$

where $r_1, \ldots, r_s \in \{0, \ldots, n-1\}$. Thus, with the notation of Proposition 1.9,

$$F(r_1, \ldots, r_s, n) = F(a_1, r_1, n) \cap \ldots \cap F(a_s, r_s, n)$$

Proposition 1.10. *A language is n-commutative if and only if it is a disjoint union of elementary n-commutative languages.*

Proof. Let $A = \{a_1, \ldots, a_s\}$, let G be a group in $\mathbf{Ab}(n)$ and let $\varphi : A^* \to G$ be a morphism. If L is recognised by φ, then $L = \varphi^{-1}(P)$ for some subset P of G. Put $\varphi(a_1) = g_1, \ldots, \varphi(a_s) = g_s$. Let $u \in A^*$ and, for $1 \leqslant i \leqslant s$, let $|u|_{a_i} \equiv r_i \bmod n$. Adopting an additive notation for G, we get

$$\varphi(u) = \sum_{1 \leqslant i \leqslant s} |u|_{a_i} g_i = \sum_{1 \leqslant i \leqslant s} r_i g_i$$

Therefore $u \in L$ if and only if $\sum_{1 \leqslant i \leqslant s} r_i g_i \in P$ and hence

$$L = \bigcup_{(r_1, \ldots, r_s) \in E} F(r_1, \ldots, r_s, n)$$

where $E = \{(r_1, \ldots, r_s) \mid \sum_{1 \leqslant i \leqslant s} r_i g_i \in P\}$. This concludes the proof, since the languages $F(r_1, \ldots, r_s, n)$ are clearly pairwise disjoint. □

Languages recognised by commutative monoids

Let $\mathcal{C}om$ be the variety of languages corresponding to the variety of commutative monoids. The following result is now a consequence of Propositions 1.8 and 1.10.

Proposition 1.11. *For each alphabet A, $\mathcal{C}om(A^*)$ is the Boolean algebra generated by the languages of the form $L(a, k)$, where $a \in A$ and $k > 0$, and $F(a, k, n)$, where $a \in A$ and $0 \leqslant k < n$.*

1.3 $\mathcal{R}$-trivial and $\mathcal{L}$-trivial languages

We study in this section the languages whose syntactic monoids are $\mathcal{R}$-trivial or $\mathcal{L}$-trivial. Surprinsingly, the corresponding results for $\mathcal{J}$-trivial and $\mathcal{H}$-trivial monoids are noticeably more difficult and will be treated in separate chapters (Chapters VI and VII).

Let $\mathcal{A} = (Q, A, \cdot)$ be a complete deterministic automaton. We say that $\mathcal{A}$ is *extensive* if there is a partial ordering $\leqslant$ on Q such that for every $q \in Q$ and for every $a \in A$, one has $q \leqslant q \cdot a$. It is important to note that although Q is equipped with a partial order, we do not require $\mathcal{A}$ to be an ordered automaton. In other words, we do not assume that the action of each letter is order preserving.

Proposition 1.12. *The transition monoid of an extensive automaton is $\mathcal{R}$-trivial.*

Proof. Let $\mathcal{A} = (Q, A, \cdot)$ be an extensive automaton and let u, v, x, y be words of A^*. Suppose that ux and v on the one hand, and vy and u on the other hand, have the same action on Q. It then follows, for every $q \in Q$, that $q \cdot u \leqslant q \cdot ux = q \cdot v$ and $q \cdot v \leqslant q \cdot vy = q \cdot u$, whence $q \cdot u = q \cdot v$ and therefore u and v have the same action on Q. It follows from this that the transition monoid of $\mathcal{A}$ is $\mathcal{R}$-trivial. □

One can deduce from Proposition 1.13 a first characterisation of languages recognised by an $\mathcal{R}$-trivial monoid.

Proposition 1.13. *A language is recognised by an $\mathcal{R}$-trivial monoid if and only if it is recognised by an extensive automaton.*

Proof. If a language is recognised by an extensive automaton, it is recognised by the transition monoid of this automaton by Proposition IV.3.19. Now this monoid is $\mathcal{R}$-trivial by Proposition 1.12.

Conversely, let L be a language of A^* recognised by an $\mathcal{R}$-trivial monoid. Then there exist a morphism $\eta : A^* \to M$ and a subset P of M such that $\eta^{-1}(P) = L$. We have seen in Section IV.3 that the automaton $\mathcal{A} = (M, A, \cdot, 1, P)$, defined by $m \cdot a = m\eta(a)$, recognises L. Since $m \geqslant_{\mathcal{R}} m\eta(a)$, the order $\geqslant_{\mathcal{R}}$ makes $\mathcal{A}$ an extensive automaton. Thus L is recognised by an extensive automaton. □

We now describe the variety of languages $\mathcal{R}$ corresponding to **R**.

Theorem 1.14. *For each alphabet A, $\mathcal{R}(A^*)$ consists of the languages which are disjoint unions of languages of the form $A_0^* a_1 A_1^* a_2 \cdots a_k A_k^*$, where $k \geqslant 0$, $a_1, \ldots, a_k \in A$, $A_k \subseteq A$ and, for $0 \leqslant i \leqslant k-1$, A_i is a subset of $A - \{a_{i+1}\}$.*

Proof. Let $L = A_0^* a_1 A_1^* a_2 \cdots a_k A_k^*$, with $k \geqslant 0$, $a_1, \ldots, a_k \in A$, $A_k \subseteq A$ and, for $0 \leqslant i \leqslant k-1$, A_i is a subset of $A - \{a_{i+1}\}$. Let us set, for $0 \leqslant i \leqslant k$, $A'_i = A - (A_i \cup \{a_{i+1}\})$. Then the automaton represented in Figure 1.1 recognises L.

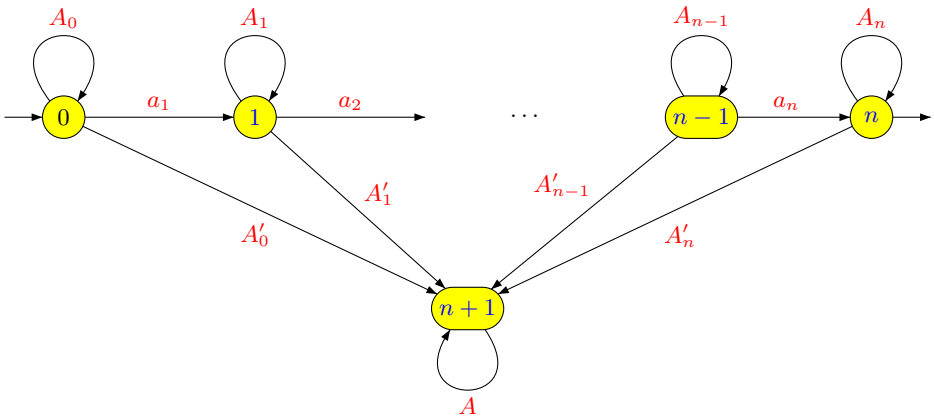


Figure 1.1. An automaton recognising L.

Since this automaton is extensive for the natural order on the integers, it follows by Proposition 1.13 that L belongs to $\mathcal{R}(A^*)$.

Conversely, let $L \in \mathcal{R}(A^*)$. By Proposition 1.13, L is recognised by an extensive automaton $\mathcal{A} = (Q, A, \cdot, q_-, F)$. Let S be the set of sequences of the form $(q_0, a_1, \ldots, a_k, q_k)$ such that $q_0 < q_1 < \cdots < q_n$, $q_0 = q_-$, $q_0 \cdot a_1 = q_1, \ldots, q_{k-1} \cdot a_k = q_k$ and $q_k \in F$. Setting, for $q \in Q$, $A_q = \{a \in A \mid q \cdot a = q\}$, we claim that

$$L = \bigcup_{(q_0, a_1, \ldots, a_k, q_k) \in S} A_{q_0}^* a_1 A_{q_1}^* \cdots A_{q_{k-1}}^* a_k A_{q_k}^*. \tag{1.1}$$

Note also that this union is disjoint since $\mathcal{A}$ is deterministic. Let K be the right-hand side of (1.1). If $u \in K$, there is a sequence $(q_0, a_1, \ldots, a_k, q_k) \in S$ and a factorisation $u = u_0 a_1 u_1 \cdots a_k u_k$ such that $u_0 \in A^*_{q_0}$, $\ldots$, $u_k \in A^*_{q_k}$. It follows that $q_0 \cdot u_0 = q_0$, $q_0 \cdot a_1 = q_1$, $\ldots$, $q_{k-1} \cdot a_k = q_k$ and $q_k \cdot u_k = q_k$. Therefore, $q_0 \cdot u = q_k$ and thus $u \in L$ since $q_0 = q_-$ and $q_k \in F$.

Conversely, let $u \in L$. Since $\mathcal{A}$ is extensive, the successful path with label u visits successively an increasing sequence of states $q_- = q_0 < q_1 < \ldots < q_k \in F$. This gives a factorisation $u = u_0 a_1 u_1 \cdots a_k u_k$, such that $q_0 \cdot u_0 = q_0$, $q_0 \cdot a_1 = q_1$, $\ldots$, $q_{k-1} \cdot a_k = q_k$ and $q_k \cdot u_k = q_k$. Consequently, u belongs to K. This proves the claim and the theorem. □

A dual result holds for the variety of languages $\mathcal{L}$ corresponding to $\mathbf{L}$.

Theorem 1.15. *For each alphabet A, $\mathcal{L}(A^*)$ consists of the languages which are disjoint unions of languages of the form $A_0^* a_1 A_1^* a_2 \cdots a_k A_k^*$, where $k \geqslant 0$, $a_1, \ldots, a_k \in A$, $A_0 \subseteq A$ and, for $1 \leqslant i \leqslant k$, A_i is a subset of $A - \{a_i\}$.*

We conclude this section with a representation theorem for $\mathcal{R}$-trivial monoids. Recall that a function f from an ordered set $(E, \leqslant)$ to itself is *extensive* if for all $x \in E$, $x \leqslant f(x)$. Let $\mathcal{E}_n$ denote the submonoid of $\mathcal{T}_n$ consisting of all extensive functions from $\{1, \ldots, n\}$ to itself.

Proposition 1.16. *For every $n > 0$, the monoid $\mathcal{E}_n$ is $\mathcal{R}$-trivial.*

Proof. Let $f, g \in \mathcal{E}_n$ such that $f \mathcal{R} g$. There exist $a, b \in \mathcal{E}_n$ such that $fa = g$ and $gb = f$. Let $q \in \{1, \ldots, n\}$. Since a is extensive one has $q \cdot f \leqslant q \cdot fa = q \cdot g$ and likewise $q \cdot g \leqslant q \cdot gb = q \cdot f$. It follows that $f = g$ and thus $\mathcal{E}_n$ is $\mathcal{R}$-trivial. □

Theorem 1.17. *A finite monoid is $\mathcal{R}$-trivial if and only if it is a submonoid of $\mathcal{E}_n$ for some $n > 0$.*

Proof. If M is $\mathcal{R}$-trivial, the relation $\geqslant_{\mathcal{R}}$ is a partial ordering. It follows that, for every $m \in M$, the right translation $\rho_m : M \to M$ defined by $\rho_m(x) = xm$ is extensive for this order. We have seen in Proposition II.4.25 that the function $m \mapsto \rho_m$ is an injective morphism from M to $\mathcal{T}(M)$. Therefore, M is a submonoid of $\mathcal{E}_n$ with $n = |M|$.

Conversely, suppose that M is a submonoid of $\mathcal{E}_n$. Then $\mathcal{E}_n$ is $\mathcal{R}$-trivial by Proposition 1.16 and M is also $\mathcal{R}$-trivial since the $\mathcal{R}$-trivial monoids form a variety of monoids. □

1.4 Some examples of +-varieties

Varieties corresponding to N, $\mathbf{N}^+$ and $\mathbf{N}^-$

Let $\mathcal{N}$ [$\mathcal{N}^+$, $\mathcal{N}^-$] denote the +-variety corresponding to $\mathbf{N}$ [$\mathbf{N}^+$, $\mathbf{N}^-$]. Recall that a subset F of a set E is *cofinite* if the complement of F in E is finite.

Proposition 1.18. *For each alphabet A,*

(1) *$\mathcal{N}^+(A^+)$ consists of the empty set and of the cofinite languages of A^+,*
(2) *$\mathcal{N}^-(A^+)$ consists of A^+ and of the finite languages of A^+,*
(3) *$\mathcal{N}(A^+)$ is the set of finite or cofinite languages of A^+.*

Proof. (1). Denote by $\varphi : A^+ \to S$ the syntactic morphism of L. If L is empty, S is trivial, and thus belongs to $\mathbf{N}^+$. If L is a cofinite subset of A^+, there exists an integer n such that L contains all the words of length $\geqslant n$. If u is such a word, we have $xuy \in L$ for each $x, y \in A^*$, thereby showing that all the words of A^+ of length $\geqslant n$ are syntactically equivalent and have the same image e under φ. Thus S is nilpotent by Proposition XI.4.14. Let now $s \in S$ and let $v \in \varphi^{-1}(s)$. The formal implication

$$(xvy \in L \Rightarrow xuy \in L)$$

shows that $v \leqslant_L u$, whence $s \leqslant e$ in S. Therefore $S \in \mathbf{N}^+$.

Conversely, let $(S, \leqslant) \in \mathbf{N}^+$, I an upper set of S and let $\varphi : A^+ \to S$ be a morphism of semigroups. If I is empty, $\varphi^{-1}(I)$ is also empty. Otherwise, I contains necessarily 0, since 0 is maximal for $\leqslant$. Let u be a word of length greater than or equal to $|S|$. By Proposition XI.4.14, $\varphi(u) = 0$ and hence $\varphi(u) \in I$. Therefore $\varphi^{-1}(I)$ is cofinite.

(2) follows from (1) by taking the complement.

(3) What precedes shows that the syntactic semigroup of a finite or cofinite subset is a nilpotent semigroup. To prove the converse, consider a nilpotent nonempty semigroup S. Let P be a subset of S and let $\varphi : A^+ \to S$ be a morphism of semigroups. Then 0 belongs either to P, or to $S - P$ and the argument above shows that $\varphi^{-1}(P)$ is either finite or cofinite. □

Varieties corresponding to $\ell\mathbf{1}$, $\mathrm{r}\mathbf{1}$ and $\mathbb{L}\mathbf{1}$

Let $\mathcal{LI}$ [$\ell\mathcal{I}$, $\mathrm{r}\mathcal{I}$] denote the +-variety corresponding to $\mathbb{L}\mathbf{1}$ [$\ell\mathbf{1}$, $\mathrm{r}\mathbf{1}$].

Proposition 1.19. *For each alphabet A, $\ell\mathcal{I}(A^+)$ [$\mathrm{r}\mathcal{I}(A^+)$] is the set of languages of the form $FA^* \cup G$ [$A^*F \cup G$] where F and G are finite languages of A^+. It is also the Boolean algebra generated by the languages of the form uA^* [A^*u].*

Proof. Let L be a language of the form $FA^* \cup G$, where F and G are finite languages of A^+. We claim that the syntactic semigroup of L belongs to $\ell\mathbf{1}$. Since the syntactic semigroup of G is nilpotent by Proposition 1.18, it suffices to consider the syntactic semigroup S of the language FA^*. Let n be the maximum length of the words of F and let u be a word of length greater than or equal to n. Then for every word $v \in A^*$, one has $uv \sim_{FA^*} u$, since the three conditions $xuvy \in FA^*$, $xu \in FA^*$ and $xuy \in FA^*$ are clearly equivalent. It follows that $ts = t$ for every $t \in S^n$ and by Proposition XI.4.15, S belongs to $\ell\mathbf{1}$.

Conversely, let L be a language recognised by a semigroup $S \in \ell\mathbf{1}$ and let $n = |S|$. Then there exists a morphism $\varphi : A^+ \to S$ such that $L = \varphi^{-1}(\varphi(L))$. Let u be a word of L of length greater than or equal to n. Let us write $u = vs$ with $|v| = n$ and $s \in A^*$. By Proposition XI.4.15 (4), one has $\varphi(u) = \varphi(v)\varphi(s) = \varphi(vA^*)$. It follows that $L = FA^* \cup G$, where F is the set of words of L of length n and G is the set of words of L of length less than n.

The second part of the statement is easy to prove. First, we know that $\ell\mathcal{I}(A^+)$ is a Boolean algebra and by the first part of the statement, it contains the languages of the form uA^*, where $u \in A^+$. Furthermore, the formula

$$\{u\} = uA^* \setminus \Big(\bigcup_{a \in A} uaA^*\Big)$$

shows that the Boolean algebra generated by the languages of the form uA^* also contains the languages of the form $\{u\}$. Therefore, it contains the languages of the form $FA^* \cup G$, with F and G finite. □

Proposition 1.20. *For each alphabet A, $\mathcal{LI}(A^+)$ is the set of languages of the form $FA^*G \cup H$ where F, G and H are finite languages of A^+. It is also the Boolean algebra generated by the languages of the form uA^* or A^*u, where $u \in A^+$.*

Proof. Let L be a language of the form $FA^*G \cup H$, where F, G and H are finite languages of A^+. We claim that the syntactic semigroup of L belongs to $\mathbb{L}\mathbf{1}$. Since the syntactic semigroup of H is nilpotent by Proposition 1.18, it suffices to consider the syntactic semigroup S of the language FA^*G. Let n be the maximum length of the words of F and G and let u be a word of length greater than or equal to n. Then for every word $v \in A^*$, one has $uvu \sim_{FA^*G} u$, since the conditions $xuvuy \in FA^*G$ and $xuy \in FA^*G$ are clearly equivalent. It follows that $tst = t$ for every $t \in S^n$ and by Proposition XI.4.17, S belongs to $\mathbb{L}\mathbf{1}$.

Conversely, let L be a language recognised by a semigroup $S \in \mathbb{L}\mathbf{1}$ and let $n = |S|$. Then there exists a morphism $\varphi : A^+ \to S$ such that $L = \varphi^{-1}(\varphi(L))$. Let u be a word of L of length greater than or equal to $2n$. Let us write $u = pvs$ with $|p| = |s| = n$. By Proposition XI.4.17, one has $\varphi(u) = \varphi(ps) = \varphi(pA^*s)$ and hence $pA^*s \subseteq L$. It follows that L is a finite union of languages of the form pA^*s, with $|p| = |s| = n$ and of a set of words of length less than $2n$.

For the second part of the statement, we know that $\mathcal{LI}(A^+)$ is a Boolean algebra and by the first part of the statement, it contains the languages of the form uA^*, A^*u, where $u \in A^+$. Furthermore, the formula

$$\{u\} = (uA^* \cap A^*u) \setminus (\bigcup_{a \in A} uaA^*)$$

shows that the Boolean algebra generated by the languages of the form uA^* or A^*u also contains the languages of the form $\{u\}$. Therefore, it contains the languages of the form $FA^*G \cup H$, with F, G and H finite. □

1.5 Cyclic and strongly cyclic languages

Cyclic languages

A language is *cyclic* if it is closed under conjugation, power and root. Thus L is cyclic if and only if, for all $u, v \in A^*$ and $n > 0$,

(1) $u^n \in L$ if and only if $u \in L$,

(2) $uv \in L$ if and only if $vu \in L$.

Let $L \in \mathrm{Rec}(A^*)$, let $\varphi : A^* \to M$ be a surjective morphism fully recognising L and let $P = \eta(L)$. A direct translation of the definition shows that a recognisable language is cyclic if and only if it satisfies, for all $x, y \in M$ and $n > 0$,

$$(xy \in P \iff yx \in P) \quad \text{and} \quad (x^n \in P \iff x \in P)$$

With the notation of Section XII.1, we get:

Proposition 1.21. *A recognisable language is cyclic if and only if it satisfies the identities* $xy \leftrightarrow yx$ *and* $x^\omega \leftrightarrow x$.

The closure properties of cyclic languages are summarised as follows.

Proposition 1.22. *Recognisable cyclic languages are closed under inverses of morphisms and under Boolean operations but not under quotients.*

Proof. The first part of the statement follows from Theorem XIII.1.2 and Proposition 1.21. Furthermore, the language $L = (abc)^* \cup (bca)^* \cup (cab)^*$ is cyclic, but its quotient $a^{-1}L = bc(abc)^*$ is not cyclic. □

Here is a useful property of the groups of a monoid fully recognising a cyclic language. This property holds of course for the syntactic monoid but we shall need this slightly more general result in the proof of Theorem 1.32.

Proposition 1.23. *Let* M *be a monoid fully recognising a recognisable cyclic language* L *and let* H *be a group of* M. *Let also* P *be the image of* L *in* M. *Then either* H *and* P *are disjoint or* H *is a subset of* P. *In the latter case, all the groups in the* $\mathcal{D}$*-class of* H *are also contained in* P.

Proof. Suppose that $P \cap H$ contains some element x. Then x^ω is the identity of H and it belongs to P by Condition (1) of the definition of a cyclic language. Now, for any element $h \in H$, we get $h^\omega = x^\omega$ and thus $h \in P$ by (1). Therefore H is contained in P.

Suppose now that P contains H and let H' be another group in the same $\mathcal{D}$-class. Let e and e' be the idempotents of H and H', respectively. By Proposition V.2.22, e and e' are conjugate and it follows from Condition (2) that e' is in P. Now, by the first part of the proposition, H' is contained in P. □

Corollary 1.24. *The syntactic monoid of a recognisable cyclic language has a zero.*

Proof. Let I be the minimal ideal of M. By Proposition V.4.37, I is a regular simple semigroup and by Proposition 1.23, one has either $I \subseteq P$ or $I \cap P = \emptyset$. It follows that any two elements s and t of I satisfy $s \sim_P t$. Indeed, for all $x, y \in M$, one gets $xsy, xty \in I$ and thus the conditions $xsy \in P$ and $xty \in P$ are equivalent. Since M is the syntactic monoid of P, one has $s = t$, which means that I has a unique element and this element is necessarily a zero of M. □

Example 1.1. Consider the language $L = b(b + ab)^*(1 + a) + ab(b + ab)^*$. The minimal automaton of this language is represented in Figure 1.2.

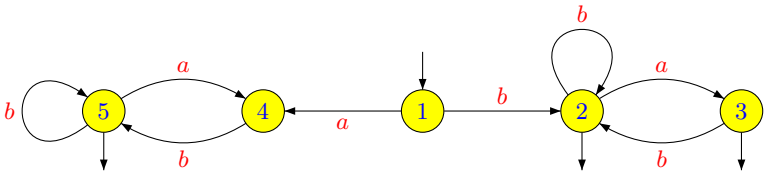


Figure 1.2. The minimal automaton of L.

The syntactic monoid of L is the monoid with zero presented by the relations

$$b^2 = b \qquad a^2 = 0 \qquad aba = a \qquad bab = b$$

Its $\mathcal{J}$-class structure is represented below. The syntactic image of L is $\{b, ab, ba\}$. It follows that L is cyclic, a property that is not so easy to see from the regular expression representing L.

^{*1}

a	$^*ab$
$^*ba$	$^*b$

$^*a^2$

Strongly cyclic languages

Let $\mathcal{A} = (Q, A, \cdot)$ be a finite (possibly incomplete) deterministic automaton. A word u *stabilises* a subset P of Q if $P \cdot u = P$. Given a subset P of Q, we let $\mathrm{Stab}(P)$ denote the set of all words that stabilise P. We also let $\mathrm{Stab}(\mathcal{A})$ denote the set of words which stabilise at least one nonempty subset P of Q: it is by definition the language that stabilises $\mathcal{A}$.

Proposition 1.25. *The language* $\mathrm{Stab}(\mathcal{A})$ *is the set of words* u *such that, for some state* q *of* $\mathcal{A}$, $q \cdot u^n$ *is defined for all* $n > 0$.

Proof. If $u \in \mathrm{Stab}(\mathcal{A})$, then u stabilises some nonempty subset P of Q. Therefore for each state $q \in P$, $q \cdot u^n$ is defined for all $n > 0$.

Conversely, suppose that for some state q of $\mathcal{A}$, $q \cdot u^n$ is defined for all $n > 0$. Then there exist two integers $k < m$ such that $q \cdot u^k = q \cdot u^m$. It follows that u stabilises the set $\{q \cdot u^i \mid k \leqslant i \leqslant m\}$. □

Example 1.2. If $\mathcal{A}$ is the automaton represented in Figure 1.3, then

$$\mathrm{Stab}(\{1\}) = (b + aa)^*, \ \mathrm{Stab}(\{2\}) = (ab^*a)^*, \ \mathrm{Stab}(\{1, 2\}) = a^*$$

and $\mathrm{Stab}(\mathcal{A}) = (b + aa)^* + (ab^*a)^* + a^*$.

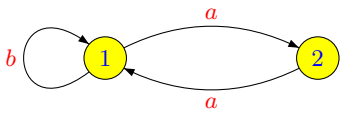


Figure 1.3. The automaton $\mathcal{A}$.

A language is *strongly cyclic* if it stabilises some finite deterministic automaton.

Proposition 1.26. *Let L be a nonfull recognisable language. The following conditions are equivalent:*

(1) *L is strongly cyclic,*
(2) *there is a morphism φ from A^* onto a finite monoid T with zero such that $L = \varphi^{-1}(\{t \in T \mid t^\omega \neq 0\})$,*
(3) *the syntactic monoid M of L has a zero and its syntactic image is the set of all elements $s \in M$ such that $s^\omega \neq 0$.*

Proof. (1) implies (2). If L is strongly cyclic, it stabilises a trimmed automaton $\mathcal{A}$. Let T be the transition monoid of $\mathcal{A}$, let k be the exponent of T and let $\varphi : A^* \to T$ be the natural morphism. Let $u \in A^*$ and let $t = \varphi(u)$. If $u \notin L$ then by Proposition 1.25, there exists for each state q an integer n_q such that $q \cdot u^{n_q}$ is undefined. Consequently, if n is larger than all the n_q, then $q \cdot u^n$ is undefined for every q. Therefore $t^n = t^\omega$ and this element is a zero of T. If now $u \in L$, then for some state q of $\mathcal{A}$, $q \cdot u^n$ is defined for all $n > 0$ and thus t^ω is not a zero of T. This proves (2).

(2) implies (3). Suppose that (2) holds and let $R = \{t \in T \mid t^\omega \neq 0\}$. Let $\eta : A^* \to M$ be the syntactic morphism of L and let P be its syntactic image. By Proposition IV.4.25, there is a surjective morphism $\pi : T \to M$ such that $\eta = \pi \circ \varphi$ and $R = \pi^{-1}(P)$. It follows that M has a zero. Furthermore, this zero is not in P, otherwise R would contain the zero of T. Let now u be a word of A^*. Let $t = \varphi(u)$ and $s = \pi(t)$. If $t^\omega = 0$, then $s^\omega = 0$. If $t^\omega \neq 0$, then $t^\omega \in R$. It follows that $s^\omega \in P$ and thus $s^\omega \neq 0$. Consequently, the conditions $t^\omega \neq 0$ and $s^\omega \neq 0$ are equivalent, which proves (3).

(3) implies (1). Suppose now that M has a zero and that P is the set of all elements $s \in M$ such that $s^\omega \neq 0$. We modify the construction given in the proof of Proposition IV.3.20. We take $\mathcal{A} = (M - 0, A, \cdot, 1, P)$ where the action is defined as follows

$$s \cdot a = \begin{cases} s\eta(a) & \text{if } s\eta(a) \neq 0 \\ \text{undefined} & \text{otherwise} \end{cases}$$

We claim that L stabilises $\mathcal{A}$. Let $u \in A^*$ and $s = \eta(u)$. If $u \in L$, then $s \cdot u^n = s^{n+1}$. This element is nonzero, for otherwise one would have $(s^{n+1})^\omega = s^\omega = 0$ which is not possible by the definition of P. If now $u \notin L$, then there is an integer n such that $s^n = 0$ and thus the transition $q \cdot u^n$ is undefined for all $q \in M - 0$ which proves the claim. □

Example 1.3. Let $L = (b + aa)^* + (ab^*a)^* + a^*$ be the strongly cyclic language considered in Example 1.2. The syntactic monoid of L is the monoid with zero presented by the relations $aa = 1$, $bb = b$, $abab = 0$ and $bab = 0$. Its transition table and its $\mathcal{J}$-class structure are represented below. The syntactic image of L is $\{1, a, b, aba\}$.

		1	2	3	4	5	6
$*$	1	1	2	3	4	5	6
	a	3	5	1	6	2	4
$*$	b	4	2	2	4	0	0
	ab	2	0	4	0	2	4
	ba	6	5	5	6	0	0
$*$	aba	5	0	6	0	5	6
$*$	bab	0	0	0	0	0	0

$*1$ a

$*aba$	ab
ba	$*b$

$*0$

It remains to give the equational description of strongly cyclic languages.

Proposition 1.27. *A recognisable language is strongly cyclic if and only if it satisfies the identities $ux^\omega v \to x^\omega$ and $x^\omega \leftrightarrow x$.*

Let us first prove a technical lemma.

Lemma 1.28. *A language which satisfies the identities $ux^\omega v \to x^\omega$ and $x^\omega \leftrightarrow x$ also satisfies the identities $xy \leftrightarrow yx$ and $0 \leqslant x$.*

Proof. Let L be a language which satisfies the identities $ux^\omega v \to x^\omega$ and $x^\omega \leftrightarrow x$. The main trick to get the identity $xy \leftrightarrow yx$ is hidden in the fact that $(xy)^\omega$ and $(yx)^\omega$ are conjugate. More precisely, the derivation

$$\begin{aligned} xy \leftrightarrow (xy)^\omega = (xy)^\omega(xy)^\omega = (xy)^{\omega-1}xy(xy)^{\omega-1}xy \\ = ((xy)^{\omega-1}x)(yx)^\omega y \to (yx)^\omega \leftrightarrow yx \end{aligned}$$

shows that $xy \to yx$ and the opposite identity $yx \to xy$ follows by symmetry.

Theorem XII.4.17 shows that L satisfies the identity $0 \leqslant x$ if and only if it is nondense or full. Let n be the exponent of L. Suppose that L is not full and let us prove that it is nondense. Let $x \notin L$. We claim that for all $u, v \in A^*$, $ux^n v \notin L$. Indeed, if $ux^n v \in L$, the identity $ux^\omega v \to x^\omega$ gives $x^n \in L$ and the identity $x^\omega \leftrightarrow x$ gives $x \in L$, a contradiction. This proves the claim and the lemma. □

Let us now come back to the proof of Proposition 1.27.

Proof. Let L be a recognisable strongly cyclic language, let M be its syntactic monoid and let P be its syntactic image. Lemma 1.28 shows that M has a zero. Observing that $x^\omega = (x^\omega)^\omega$, one gets

$$x \in P \iff x^\omega \neq 0 \iff (x^\omega)^\omega \neq 0 \iff x^\omega \in P$$

which proves that L satisfies the identity $x^\omega \leftrightarrow x$. Finally, if $ux^\omega v \in P$, then $(ux^\omega v)^\omega \neq 0$. Therefore $x^\omega \neq 0$, whence $x \in P$ and finally $x^\omega \in P$ since $x^\omega \leftrightarrow x$. Thus L satisfies the identity $ux^\omega v \to x^\omega$.

Conversely, suppose that L satisfies the two identities of the statement. By Lemma 1.28, M has a zero and $0 \notin P$. By Proposition 1.26, it suffices to prove that $x \in P$ if and only if $x^\omega \neq 0$. First, if $x \in P$, then $x^\omega \in P$ and since $0 \notin P$, one has $x^\omega \neq 0$. Now, if $x^\omega \neq 0$, then $ux^\omega v \in P$ for some $u, v \in M$ (since x^ω is not equivalent to 0 in the syntactic equivalence of P). It follows $x^\omega \in P$ by the first identity and $x \in P$ by the second one. □

The next corollaries allows a precise comparison between cyclic and strongly cyclic languages.

Corollary 1.29. *Any recognisable strongly cyclic language is cyclic.*

Proof. It suffices to prove that a recognisable strongly cyclic language also satisfies the identity $xy \leftrightarrow yx$. Indeed, for each $x, y \in M$, $(xy)^\omega$ and $(yx)^\omega$ are two conjugate idempotents. It follows that $(xy)^\omega \neq 0$ if and only if $(yx)^\omega \neq 0$. □

Corollary 1.30. *Let L be a recognisable cyclic language. Let M be its syntactic monoid and let P be its syntactic image. Then L is strongly cyclic if and only if for all idempotent e, f of M, the conditions $e \in P$ and $e \leqslant_{\mathcal{J}} f$ imply $f \in P$.*

The proof is again a typical exercise on identities.

Proof. Let $\varphi : A^* \to M$ be the syntactic morphism of L. Suppose that L is strongly cyclic and let e, f be two idempotents of M such that $e \leqslant_{\mathcal{J}} f$. Also let $s, t \in M$ be such that $e = sft$. Lifting up to A^*, one can find words x, y, u, v such that $\varphi(x) = e$, $\varphi(y) = f$, $\varphi(u) = s$ and $\varphi(v) = t$. By Proposition 1.27, L satisfies the identities (a) $ux^\omega v \to x^\omega$ and (b) $x^\omega \leftrightarrow x$. Now $\widehat{\varphi}(uy^\omega v)^\omega = sft = e \in P$, and thus by (b) $\widehat{\varphi}(uy^\omega v) \in P$. Since $\widehat{\varphi}(uy^\omega v) = \widehat{\varphi}(u)\widehat{\varphi}(y)^\omega\widehat{\varphi}(v)$, it follows by (a) that $\widehat{\varphi}(y)^\omega \in P$ and thus $f \in P$.

In the opposite direction, suppose that for all idempotents e, f of M, the conditions $e \in P$ and $e \leqslant_{\mathcal{J}} f$ imply $f \in P$. Since L is cyclic it satisfies the identity $x^\omega \leftrightarrow x$ by Proposition 1.21. We claim that it also satisfies the identity $ux^\omega v \to x^\omega$. First, $\widehat{\varphi}(ux^\omega v) \in P$ implies $\widehat{\varphi}(ux^\omega v)^\omega \in P$ since $x^\omega \leftrightarrow x$. Furthermore, since $\widehat{\varphi}(ux^\omega v)^\omega \leqslant_{\mathcal{J}} \widehat{\varphi}(x^\omega)$, one also has $\widehat{\varphi}(x^\omega) \in P$, which finally gives $\widehat{\varphi}(x) \in P$ since $x^\omega \leftrightarrow x$. □

Proposition 1.31. *Recognisable strongly cyclic languages are closed under inverses of morphisms, finite intersection and finite union but not under quotients.*

Proof. Theorem XIII.1.2 and Proposition 1.27 show that strongly cyclic languages form a positive stream of languages. It remains to show that there are not closed under quotients. The language $L = (b + aa)^* + (ab^*a)^* + a^*$ of Example 1.3 is strongly cyclic. However, its quotient $b^{-1}L = (b + aa)^*$ is not strongly cyclic, since $aa \in (b + aa)^*$ but $a \notin (b + aa)^*$. □

We now give the connection between cyclic and strongly cyclic languages.

Theorem 1.32. *A recognisable language is cyclic if and only if it is a Boolean combination of recognisable strongly cyclic languages.*

Proof. Propositions 1.21 and 1.27 and the remark preceeding Proposition 1.31 show that a Boolean combination of recognisable strongly cyclic language is cyclic.

We now show that any cyclic language L is a Boolean combination of recognisable strongly cyclic languages. The proof is by induction on the size of a monoid with zero fully recognising L. Such a monoid always exists, since by Corollary 1.24, the syntactic monoid of L has a zero.

Let T be a monoid with zero, let $\varphi : A^* \to T$ be a surjective morphism recognising L and let $P = \varphi(L)$. We may also assume that $0 \notin P$. Otherwise,

observing that $0 \notin P^c$, it suffices to prove the result for L^c, which is also cyclic and fully recognised by φ.

If T is trivial, then $1 = 0$ and P is empty. Thus L is necessarily the empty language, which is strongly cyclic since it stabilises the empty automaton. Suppose now that T is nontrivial. Let us set

$$R = \{t \in T \mid t^\omega \neq 0\},\ S = R - P,\ U = \varphi^{-1}(R) \text{ and } V = U - L$$

Proposition 1.26 shows that U is strongly cyclic. Moreover P is a subset of R and thus L is contained in U. Also note that $V = \varphi^{-1}(S)$. Let D be a 0-minimal $\mathcal{J}$-class of T and let $s \in D$. Then s is a nonzero element such that $t \leqslant_{\mathcal{J}} s$ implies $t \mathcal{J} s$ or $t = 0$.

Suppose first that $P \cap D = \emptyset$. We claim that $s \sim_P 0$. Indeed, let $x, y \in T$. Since $x0y = 0$ and $0 \notin P$, it suffices to prove that $xsy \notin P$. But this is clear, since xsy belongs to $D \cup \{0\}$, which is disjoint from P. Therefore L is fully recognised by $T/\!\sim_P$, a strict quotient of T, and the induction hypothesis gives the result.

Suppose now that $P \cap D \neq \emptyset$ and let $t \in P \cap D$. Then $t^\omega \in P$ since L is cyclic. In particular $t^\omega \neq 0$ and thus $t^\omega \in P \cap D$. We claim that $t \sim_S 0$. Again, as $0 \notin S$, it suffices to prove that for all $x, y \in T$, $xty \notin S$. We consider two cases. If $(xty)^\omega = 0$, then $xty \notin R$ by the definition of R. Now, if $(xty)^\omega \neq 0$, then $(xty)^\omega \mathcal{J} t^\omega$ since D is a 0-minimal $\mathcal{J}$-class. But since $t^\omega \in P$, we also have $(xty)^\omega \in P$ by Proposition 1.23.

It follows that the language V is fully recognised by $T/\!\sim_S$, a strict quotient of T. Since V is a cyclic language, the induction hypothesis tells us that V is a Boolean combination of recognisable strongly cyclic languages and so is $L = U - V$. □

2 Exercises

Exercise 1. Consider the variety of semigroups $\mathbf{N}_k = [\![x_1 \cdots x_k = y_1 \cdots y_k]\!]$ and let $\mathcal{N}_k$ be the corresponding $+$-variety of languages.

(1) Show that a nonempty semigroup S is in $\mathbf{N}_k$ if and only if it has a zero and satisfies $S^k = 0$.

(2) Show that, for every alphabet A, $\mathcal{N}_k(A^+)$ is the Boolean algebra generated by the languages of the form $\{u\}$ where $|u| < k$.

Exercise 2. Consider the variety of semigroups $\ell\mathbf{1}_k = [\![x_1 \cdots x_k y = x_1 \cdots x_k]\!]$ and let $\ell\mathcal{I}_k$ be the corresponding $+$-variety of languages. Show that, for every alphabet A, $\ell\mathcal{I}_k(A^+)$ is the set of all languages of the form $FA^* \cup G$, where F and G are subsets of A^+ formed of words of length equal to or less than k and less than k respectively.

Exercise 3. Consider the variety of semigroups $\mathbb{L}\mathbf{1}_k = [\![x_1 \cdots x_k y x_1 \cdots x_k = x_1 \cdots x_k]\!]$ and let $\mathcal{LI}_k$ be the corresponding $+$-variety of languages. Show that, for every alphabet A, $\mathcal{LI}_k(A^+)$ is the Boolean algebra generated by the languages of the form uA^* and A^*u, where $|u| \leqslant k$.

Exercise 4. Let $\mathbf{V}$ be the variety considered in Exercise XI.8. and let $\mathcal{V}$ be the variety of languages corresponding to $\mathbf{V}$. Show that for each alphabet A, $\mathcal{V}(A^*)$ is the Boolean algebra generated by the languages of the form $B^*a_1B^*a_2B^* \cdots a_nB^*$ where $a_1, \ldots, a_n \in A \setminus B$.

3 Notes

The results of Section 1.5 are adapted from [11].

Part D

Advanced tools

Chapter XV

Polynomial closure

The *polynomial closure* $\mathrm{Pol}(\mathcal{L})$ of a class of languages $\mathcal{L}$ of A^* is the set of languages that are finite unions of marked products of the form $L_0a_1L_1\cdots a_nL_n$, where the a_i are letters and the L_i are elements of $\mathcal{L}$.

The main result of this chapter gives an equational description of $\mathrm{Pol}(\mathcal{L})$, given an equational description of $\mathcal{L}$, when $\mathcal{L}$ is a lattice of languages closed under quotient. It can be formally stated as follows:

> *If $\mathcal{L}$ is a lattice of languages closed under quotients, then Pol($\mathcal{L}$) is defined by the set of equations of the form $x^\omega \leqslant x^\omega y x^\omega$, where x, y are profinite words such that the equations $x = x^2$ and $x \leqslant y$ are satisfied by $\mathcal{L}$.*

1 Polynomial closure of a lattice of languages

Let $\mathcal{L}$ be a set of languages of A^*. An *$\mathcal{L}$-monomial of degree n* is a language of the form $L_0a_1L_1\cdots a_nL_n$, where each a_i is a letter of A and each L_i is a language of $\mathcal{L}$. An *$\mathcal{L}$-polynomial* is a finite union of $\mathcal{L}$-monomials. Its *degree* is the maximum of the degrees of its monomials. The *polynomial closure* of $\mathcal{L}$, denoted by $\mathrm{Pol}(\mathcal{L})$, is the set of all $\mathcal{L}$-polynomials.

Our main result gives an equational description of $\mathrm{Pol}(\mathcal{L})$, given an equational description of $\mathcal{L}$, when $\mathcal{L}$ is a lattice of languages closed under quotients. To state this result in a concise way, let us introduce a convenient notation. Given a set $\mathcal{R}$ of recognisable languages, let $\Sigma(\mathcal{R})$ denote the set of equations of the form $x^\omega \leqslant x^\omega y x^\omega$, where x, y are profinite words of $\widehat{A^*}$ such that the equations $x = x^2$ and $x \leqslant y$ are satisfied by $\mathcal{R}$. Note that the function mapping $\mathcal{R}$ to the class of languages satisfying $\Sigma(\mathcal{R})$ is monotonic with respect to the inclusion. We can now state our main result:

Theorem 1.1. *If $\mathcal{L}$ is a lattice of languages closed under quotients, then Pol($\mathcal{L}$) is defined by the set of equations $\Sigma(\mathcal{L})$.*

The proof is divided into several parts. We first prove in Proposition 1.2 that $\mathrm{Pol}(\mathcal{L})$ satisfies the equations of $\Sigma(\mathcal{L})$. To establish the converse of this property, we consider a language K satisfying all the equations of $\Sigma(\mathcal{L})$. We convert this property into a topological property (Proposition 1.3) and then use a compactness argument to show that K satisfies the equations of $\Sigma(\mathcal{F})$,

where $\mathcal{F}$ is a finite sublattice of $\mathcal{L}$ (Proposition 1.4). The final part of the proof consists in proving that K belongs to $\mathrm{Pol}(\mathcal{F})$. This is where the factorisation forest theorem arises, but a series of lemmas (Lemmas 1.5 to 1.10) are still necessary to explicitly find a polynomial expression for K.

Proposition 1.2. *If $\mathcal{L}$ is a lattice of languages, then $\mathrm{Pol}(\mathcal{L})$ satisfies the equations of $\Sigma(\mathcal{L})$.*

Proof. Since, by Theorem XII.2.6, the set of languages satisfying $\Sigma(\mathcal{L})$ is a lattice of languages, it suffices to prove the result for an $\mathcal{L}$-monomial. Let $L = L_0a_1L_1\cdots a_nL_n$ be an $\mathcal{L}$-monomial and let $\eta\colon A^* \to M$ be its syntactic morphism. Let, for $0 \leqslant i \leqslant n$, $\eta_i\colon A^* \to M_i$ be the syntactic morphism of L_i. Let x and y be two profinite words such that each L_i satisfies the two equations $x = x^2$ and $x \leqslant y$.

Since A^* is dense in $\widehat{A^*}$, one can find a word $x' \in A^*$ such that $r(x', x) > \max\{|M_0|, \ldots, |M_n|, |M|\}$. It follows that $\eta(x') = \widehat{\eta}(x)$ and, for $0 \leqslant i \leqslant n$, $\eta_i(x') = \widehat{\eta}_i(x)$. Similarly, one can associate with y a word $y' \in A^*$ such that $\eta(y') = \widehat{\eta}(y)$ and, for $0 \leqslant i \leqslant n$, $\eta_i(y') = \widehat{\eta}_i(y)$. It follows that each L_i satisfies the equations $x' = x'^2$ and $x' \leqslant y'$ and that L satisfies the equation $x^\omega \leqslant x^\omega y x^\omega$ if and only if it satisfies the equations $x'^\omega \leqslant x'^\omega y' x'^\omega$. In other terms, it suffices to prove the result when x and y are words.

We need to establish the relation

$$\widehat{\eta}(x^\omega) \leqslant \widehat{\eta}(x^\omega y x^\omega) \tag{1.1}$$

Let k be an integer such that $k > n$ and $\widehat{\eta}(x^\omega) = \eta(x^k)$. Since $\widehat{\eta}(x^\omega y x^\omega) = \eta(x^k y x^k)$, proving (1.1) amounts to showing that $x^k \leqslant_L x^k y x^k$. Let $u, v \in A^*$ and suppose that $ux^k v \in L$. Thus $ux^k v = u_0a_1u_1\cdots a_nu_n$, where, for $0 \leqslant i \leqslant n$, $u_i \in L_i$. Since $k > n$, one can find $h \in \{0, \ldots, n\}$, $j \in \{1, \ldots, k\}$ and $u'_h, u''_h \in A^*$ such that $u_h = u'_h x u''_h$, $ux^{j-1} = u_0a_1u_1\cdots a_hu'_h$ and $x^{k-j}v = u''_h a_{h+1}u_{h+1}\cdots a_nu_n$. Since $u_h \in L_h$ and L_h satisfies the equations $x = x^2$ and $x \leqslant y$, one has $u'_h x^{k-j+1} y x^j u''_h \in L_h$, and since

$$ux^k y x^k v = u_0a_1u_1\cdots a_h(u'_h x^{k-j+1} y x^j u''_h)a_{h+1}u_{h+1}\cdots a_nu_n$$

one gets $ux^k y x^k v \in L$. Thus $x^k \leqslant_L x^k y x^k$, which completes the proof. □

The rest of this section is devoted to showing the converse implication in Theorem 1.1. Let us introduce, for each recognisable language L of A^*, the sets

$$E_L = \left\{(x, y) \in \widehat{A^*} \times \widehat{A^*} \mid L \text{ satisfies } x = x^2 \text{ and } x \leqslant y\right\}$$
$$F_L = \left\{(x, y) \in \widehat{A^*} \times \widehat{A^*} \mid L \text{ satisfies } x^\omega \leqslant x^\omega y x^\omega\right\}.$$

Lemma XII.2.7 shows that the set E_L is clopen in $\widehat{A^*} \times \widehat{A^*}$. A similar argument, using the continuous map $\beta : \widehat{A^*} \times \widehat{A^*} \to M^2$ defined by

$$\beta(x, y) = \left(\widehat{\eta}(x^\omega y x^\omega), \widehat{\eta}(x^\omega)\right)$$

would show that F_L is clopen.

We now convert our equational conditions into a topological property. Recall that a *cover* [*open cover*] of a topological space X is a collection of subsets [open subsets] of X whose union is X.

Proposition 1.3. *Let $\mathcal{F}$ be a set of recognisable languages of A^* and let K be a recognisable language of A^*. The following conditions are equivalent:*

(1) *K satisfies the profinite equations of $\Sigma(\mathcal{F})$,*

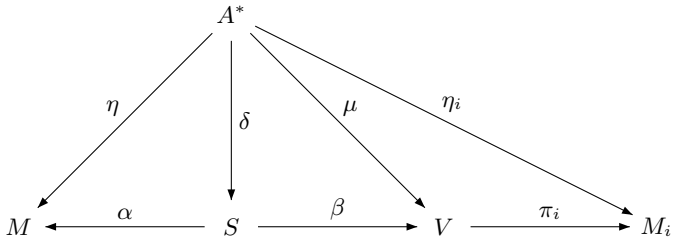
(2) *the set $\{F_K\} \cup \{E_L^c \mid L \in \mathcal{F}\}$ is an open cover of $\widehat{A^*} \times \widehat{A^*}$.*

Proof. Indeed $\mathcal{F}$ satisfies the two profinite equations $x = x^2$ and $x \leqslant y$ if and only if $(x, y) \in \bigcap_{L\in\mathcal{F}} E_L$ or, equivalently, $(x, y) \notin \bigcup_{L\in\mathcal{F}} E_L^c$. Similarly, K satisfies the equation $x^\omega \leqslant x^\omega y x^\omega$ if and only if $(x, y) \in F_K$. Now, condition (1) is equivalent to saying that $(x, y) \notin \bigcup_{L\in\mathcal{F}} E_L^c$ implies $(x, y) \in F_K$, which is another way to say that $\{F_K\} \cup \{E_L^c \mid L \in \mathcal{F}\}$ is a cover of $\widehat{A^*} \times \widehat{A^*}$. Since F_K and E_L are clopen, it is an open cover. □

Proposition 1.4. *If K satisfies the equations of $\Sigma(\mathcal{L})$, there is a finite subset $\mathcal{F}$ of $\mathcal{L}$ such that K satisfies the equations of $\Sigma(\mathcal{F})$.*

Proof. Proposition 1.3 shows that $\{F_K\} \cup \{E_L^c \mid L \in \mathcal{L}\}$ is a cover of $\widehat{A^*} \times \widehat{A^*}$. Since $\widehat{A^*}$ is compact, one can extract from this cover a finite cover, say $\{F_K\} \cup \{E_L^c \mid L \in \mathcal{F}\}$. By Proposition 1.3 again, K satisfies the profinite equations of the form $x^\omega \leqslant x^\omega y x^\omega$ such that all the languages of $\mathcal{F}$ satisfy the equations $x = x^2$ and $x \leqslant y$. □

Let K be a recognisable language satisfying all the equations of $\Sigma(\mathcal{L})$ and let $\eta : A^* \to M$ be its syntactic morphism. Let also $\mathcal{F} = \{L_1, \ldots, L_n\}$ be a finite subset of $\mathcal{L}$ as given by Proposition 1.4. For $1 \leqslant i \leqslant n$, let $\eta_i \colon A^* \to M_i$ be the syntactic morphism of L_i. Let $\mu \colon A^* \to M_1 \times \cdots \times M_n$ be the morphism defined by $\mu(u) = (\eta_1(u), \ldots, \eta_n(u))$. Finally, let $V = \mu(A^*)$ and, for $1 \leqslant i \leqslant n$, let $\pi_i : V \to M_i$ be the natural projection. We set $S = \{(\eta(u), \mu(u)) \mid u \in A^*\}$. Then S is a submonoid of $M \times V$ and the two morphisms $\alpha : S \to M$ and $\beta : S \to V$ defined by $\alpha(m, v) = m$ and $\beta(m, v) = v$ are surjective. Furthermore, the morphism $\delta : A^* \to S$ defined by $\delta(u) = (\eta(u), \mu(u))$ satisfies $\eta = \alpha \circ \delta$ and $\mu = \beta \circ \delta$. The situation is summarised in the following diagram:



We now arrive at the last step of the proof of Theorem 1.1, which consists in proving that K belongs to $\mathrm{Pol}(\mathcal{F})$.

We start with three auxiliary lemmas. The first one states that every downward closed language recognised by μ belongs to $\mathcal{L}$ and relies on the fact that $\mathcal{L}$ is a lattice of languages closed under quotients. The second one gives a key property of S and this is the only place in the proof where we use the equations of $\Sigma(\mathcal{L})$. The third one is an elementary, but useful, observation.

Lemma 1.5. *Let $t \in V$. Then the language $\mu^{-1}(\uparrow t)$ belongs to $\mathcal{L}$.*

Proof. Let $t = (t_1, \ldots, t_n)$ and let z be a word such that $\mu(z) = t$. Then $t_i = \eta_i(z)$ and $\mu^{-1}(\uparrow t) = \bigcap_{1 \leqslant i \leqslant n} \eta_i^{-1}(\uparrow t_i)$. Moreover, one gets for each $i \in \{1, \ldots, n\}$,

$$\eta_i^{-1}(\uparrow t_i) = \{x \in A^* \mid \eta_i(z) \leqslant \eta_i(x)\} = \{x \in A^* \mid z \leqslant_{L_i} x\} = \bigcap_{(u,v) \in E_i} u^{-1} L_i v^{-1}$$

where $E_i = \{(u, v) \in A^* \times A^* \mid uzv \in L_i\}$. Since L_i is recognisable, there are only finitely many quotients of the form $u^{-1}L_iv^{-1}$ and hence the intersection is finite. The result follows, since $\mathcal{L}$ is a lattice of languages closed under quotients. □

Lemma 1.6. *For every idempotent $(e, f) \in S$ and for every $(s, t) \in S$ such that $f \leqslant t$, one has $e \leqslant ese$.*

Proof. Let x and y be two words such that $\delta(x) = (e, f)$ and $\delta(y) = (s, t)$. Then $\eta(x) = e$, $\mu(x) = f$, $\eta(y) = s$ and $\mu(y) = t$ and since f is idempotent and $f \leqslant t$, $\mathcal{F}$ satisfies the equations $x = x^2$ and $x \leqslant y$. Therefore K satisfies the equation $x^\omega \leqslant x^\omega y x^\omega$. It follows that $\widehat{\eta}(x^\omega) \leqslant \widehat{\eta}(x^\omega y x^\omega)$, that is $e \leqslant ese$. □

Before we continue, let us point out a subtlety in the proof of Lemma 1.6. It looks like we have used words instead of profinite words in this proof and the reader may wonder whether one could change "profinite" to "finite" in the statement of our main result. The answer is negative for the following reason: if $\mathcal{F}$ satisfies the equations $x = x^2$ and $x \leqslant y$, it does not necessarily imply that $\mathcal{L}$ satisfies the same equations. In fact, the choice of $\mathcal{F}$ comes from the extraction of the finite cover and hence is bound to K.

We now set, for each idempotent f of V, $L(f) = \mu^{-1}(\uparrow f)$.

Lemma 1.7. *For each idempotent f of V, one has $L(1)L(f)L(1) = L(f)$.*

Proof. Since $1 \in L(1)$, one gets the inclusion $L(f) = 1L(f)1 \subseteq L(1)L(f)L(1)$. Let now $s, t \in L(1)$ and $x \in L(f)$. Then by definition, $1 \leqslant \mu(s)$, $f \leqslant \mu(x)$ and $1 \leqslant \mu(t)$. It follows that $f = 1f1 \leqslant \mu(s)\mu(x)\mu(t) = \mu(sxt)$, whence $sxt \in L(f)$. This gives the opposite inclusion $L(1)L(f)L(1) \subseteq L(f)$. □

We now come to the combinatorial argument of the proof. By Theorem II.6.40, there exists a factorisation forest F of height $\leqslant 3|S| - 1$ which is Ramseyan modulo δ. We use this fact to associate with each word x a certain language $R(x)$, defined recursively as follows:

$$R(x) = \begin{cases} L(1)xL(1) & \text{if } |x| \leqslant 1 \\ R(x_1)R(x_2) & \text{if } F(x) = (x_1, x_2) \\ R(x_1)L(f)R(x_k) & \text{if } F(x) = (x_1, \ldots, x_k), \text{ with } k \geqslant 3 \text{ and} \\ & \quad \delta(x_1) = \cdots = \delta(x_k) = (e, f) \end{cases}$$

In particular $R(1) = L(1)$, since $L(1)$ is a submonoid of A^*. The *height function* of F is the function $h : A^+ \to \mathbb{N}$ defined by

$$h(x) = \begin{cases} 0 & \text{if } x \text{ is a letter} \\ 1 + \max\{h(x_i) \mid 1 \leqslant i \leqslant k\} & \text{if } F(x) = (x_1, \ldots, x_k) \end{cases}$$

Thus $h(x)$ is the length of the longest path with origin in x in the tree of x.

Denote by $\mathcal{E}$ the finite set of languages of the form $L(f)$, where f is an idempotent of V. We know by Lemma 1.5 that $\mathcal{E}$ is a subset of $\mathcal{L}$. Let us say that an $\mathcal{E}$-monomial is in *normal form* if it is of the form $L(1)a_0L(f_1)a_1\cdots L(f_k)a_kL(1)$ where $f_1,\ldots,f_k$ are idempotents of V.

Lemma 1.8. *For each $x \in A^*$, $R(x)$ is equal to an $\mathcal{E}$-monomial in normal form of degree $\leqslant 2^{h(x)}$.*

Proof. We prove the result by induction on the length of x. The result is true if $|x| \leqslant 1$. Suppose that $|x| \geqslant 2$. If $F(x) = (x_1, x_2)$, then $R(x) = R(x_1)R(x_2)$ otherwise $R(x) = R(x_1)L(f)R(x_k)$. We treat only the latter case, since the first one is similar. By the induction hypothesis, $R(x_1)$ and $R(x_k)$ are equal to $\mathcal{E}$-monomials in normal form. It follows by Lemma 1.7 that $R(x)$ is equal to an $\mathcal{E}$-monomial in normal form, whose degree is lesser than or equal to the sum of the degrees of $R(x_1)$ and $R(x_k)$. The result now follows from the induction hypothesis, since $2^{h(x_1)} + 2^{h(x_k)} \leqslant 2^{1+\max\{h(x_1),\ldots,h(x_k)\}} \leqslant 2^{h(x)}$. □

Lemma 1.9. *For each $x \in A^*$, one has $x \in R(x)$.*

Proof. We prove the result by induction on the length of x. The result is trivial if $|x| \leqslant 1$. Suppose that $|x| \geqslant 2$. If $F(x) = (x_1, x_2)$, one has $x_1 \in R(x_1)$ and $x_2 \in R(x_2)$ by the induction hypothesis and hence $x \in R(x)$ since $R(x) = R(x_1)R(x_2)$. Suppose now that $F(x) = (x_1,\ldots,x_k)$ with $k \geqslant 3$ and $\delta(x_1) = \cdots = \delta(x_k) = (e, f)$. Then $R(x) = R(x_1)L(f)R(x_k)$. Since $x_1 \in R(x_1)$ and $x_k \in R(x_k)$ by the induction hypothesis and $\mu(x_2\cdots x_{k-1}) = f$, one gets $x_2\cdots x_{k-1} \in L(f)$ and finally $x \in R(x_1)L(f)R(x_k)$, that is, $x \in R(x)$. □

If R is a language, let us write $\eta(x) \leqslant \eta(R)$ if, for each $u \in R$, $\eta(x) \leqslant \eta(u)$.

Lemma 1.10. *For each $x \in A^*$, one has $\eta(x) \leqslant \eta(R(x))$.*

Proof. We prove the result by induction on the length of x. First, applying Lemma 1.6 with $e = f = 1$ shows that if $(s, t) \in S$ and $1 \leqslant t$, then $1 \leqslant s$. It follows that $1 \leqslant \eta(\mu^{-1}(\uparrow 1)) = \eta(L(1)) = \eta(R(1))$.

If $|x| \leqslant 1$, one gets $R(x) = L(1)xL(1)$ and $\eta(x) \leqslant \eta(L(1))\eta(x)\eta(L(1)) = \eta(R(x))$ since $1 \leqslant \eta(L(1))$. Suppose now that $|x| \geqslant 2$. If $F(x) = (x_1, x_2)$, then $R(x) = R(x_1)R(x_2)$ and by the induction hypothesis, $\eta(x_1) \leqslant \eta(R(x_1))$ and $\eta(x_2) \leqslant \eta(R(x_2))$. Therefore, $\eta(x) = \eta(x_1)\eta(x_2) \leqslant \eta(R(x_1))\eta(R(x_2)) = \eta(R(x))$. Finally, suppose that $F(x) = (x_1,\ldots,x_k)$ with $k \geqslant 3$ and $\delta(x_1) = \cdots = \delta(x_k) = (e, f)$. Then $R(x) = R(x_1)L(f)R(x_k)$. By the induction hypothesis, $e \leqslant \eta(R(x_1))$ and $e \leqslant \eta(R(x_k))$. Now, if $u \in L(f)$, one gets $f \leqslant \mu(u)$. Since $(\eta(u), \mu(u)) \in S$, it follows from Lemma 1.6 that the relation $e \leqslant e\eta(u)e$ holds in M. Finally, we get $\eta(x) = e \leqslant e\eta(L(f))e \leqslant \eta(R(x_1))\eta(L(f))\eta(R(x_k)) = \eta(R(x))$. □

We can now conclude the proof of Theorem 1.1. We claim that $K = \bigcup_{x\in K} R(x)$. The inclusion $K \subseteq \bigcup_{x\in K} R(x)$ is an immediate consequence of Lemma 1.9. To prove the opposite inclusion, consider a word $u \in R(x)$ for some $x \in K$. It follows from Lemma 1.10 that $\eta(x) \leqslant \eta(u)$. Since $\eta(x) \in \eta(K)$, one gets $\eta(u) \in \eta(K)$ and finally $u \in K$. Now, by Lemma 1.8, each language $R(x)$ is an $\mathcal{E}$-monomial of degree $\leqslant 2^{h(x)}$. Since $h(x) \leqslant 3|S| - 1$ for all x, and since $\mathcal{E}$ is finite, there are only finitely many such monomials. Therefore K is equal to

an $\mathcal{E}$-polynomial. Finally, Lemma 1.5 shows that each $\mathcal{E}$-polynomial belongs to $\mathrm{Pol}(\mathcal{L})$, and thus $K \in \mathrm{Pol}(\mathcal{L})$. □

2 A case study

As an application of this result, we establish a set of profinite equations defining the class of languages of the form $L_0a_1L_1 \cdots a_nL_n$, where each language L_i is either of the form u^* (where u is a word) or A^* (where A is the alphabet) and we prove that this class is decidable.

We have seen that the recognisable languages that are either slender or full form a lattice of languages closed under quotients, denoted by $\mathcal{S}$ in the sequel. One can show that the languages of $\mathrm{Pol}(\mathcal{S})$ are finite unions of languages of the form $L_0a_1L_1 \cdots a_nL_n$, where the a_i are letters and the L_i are languages of the form A^* or u^* for some word u. In particular, $\mathrm{Pol}(\mathcal{S})$ contains all recognisable sparse languages but it also contains the nonsparse language A^* if $|A| \geqslant 2$.

An equational description of $\mathrm{Pol}(\mathcal{S})$ was given in [18]. Let $\Sigma'(\mathcal{S})$ denote the set of equations of the form

$$(x^\omega y^\omega)^\omega \leqslant (x^\omega y^\omega)^\omega z (x^\omega y^\omega)^\omega$$

where $z \in A^*$ and $x, y \in A^+$ and $i(x) \neq i(y)$.

Theorem 2.11. *A recognisable language of A^* belongs to $Pol(\mathcal{S})$ if and only if it satisfies the equations of $\Sigma'(\mathcal{S})$.*

Chapter XVI

Relational morphisms

Relational morphisms form a powerful tool in semigroup theory. Although the study of relational morphisms can be reduced in theory to the study of morphisms, their systematic use leads to concise proofs of nontrivial results. Furthermore, they provide a natural definition of the Mal'cev product and its variants, an important tool for decomposing semigroups into simpler pieces.

1 Relational morphisms

A *relational morphism* between two semigroups S and T is a relation $\tau : S \to T$ which satisfies

(1) for every $s \in S$, $\tau(s) \neq \emptyset$,

(2) for every $s_1, s_2 \in S$, $\tau(s_1)\tau(s_2) \subseteq \tau(s_1 s_2)$

For a relational morphism between two monoids S and T, a third condition is required

(3) $1 \in \tau(1)$

The *graph* of a relational morphism τ is the subset R of $S \times T$ defined by

$$R = \{(s, t) \in S \times T \mid t \in \tau(s)\}$$

The following result is an immediate consequence of the definition of a relational morphism.

Proposition 1.1. *The graph of a relational morphism between two semigroups [monoids] S and T is a subsemigroup [submonoid] of $S \times T$.*

Let us also mention another immediate result.

Proposition 1.2. *The composition of two relational morphisms is a relational morphism.*

Examples of relational morphisms include two standard classes:

(1) morphisms,

(2) inverses of surjective morphisms.

Indeed if $\alpha : S \to T$ is a surjective morphism, then the relation $\alpha^{-1} : T \to S$ is a relational morphism. These two classes generate all relational morphisms.

More precisely, every relational morphism is the composition of a morphism and the inverse of a surjective morphism.

Proposition 1.3. *Let $\tau : S \to T$ be a relational morphism and let R be its graph. Then the projections from $S \times T$ onto S and T induce morphisms $\alpha : R \to S$ and $\beta : R \to T$ such that α is surjective and $\tau = \beta \circ \alpha^{-1}$.*

Proof. The factorisation of τ as $\beta \circ \alpha^{-1}$ is an immediate consequence of the definition. The surjectivity of α stems from the fact that, for all $s \in S$, $\tau(s)$ is nonempty. □

The factorisation $\tau = \beta \circ \alpha^{-1}$, pictured in Figure 1.1 is called the *canonical factorisation* of τ.

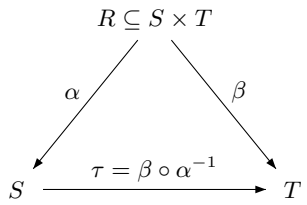


Figure 1.1. The canonical factorisation of a relational morphism.

We shall see that in most cases the properties of τ are tied to those of β (see in particular Propositions 2.5 and 3.11).

The next result extends Proposition II.3.8 to relational morphisms. We remind the reader that if τ is a relation from S to T and T' is a subset of T, then $\tau^{-1}(T') = \{s \in S \mid \tau(s) \cap T' \neq \emptyset\}$.

Proposition 1.4. *Let $\tau : S \to T$ be a relational morphism. If S' is a subsemigroup of S, then $\tau(S')$ is a subsemigroup of T. If T' is a subsemigroup of T, then $\tau^{-1}(T')$ is a subsemigroup of S.*

Proof. Let $t_1, t_2 \in \tau(S')$. Then $t_1 \in \tau(s_1)$ and $t_2 \in \tau(s_2)$ for some $s_1, s_2 \in S'$. It follows that $t_1t_2 \in \tau(s_1)\tau(s_2) \subseteq \tau(s_1s_2) \subseteq \tau(S')$ and therefore $\tau(S')$ is a subsemigroup of T.

Let $s_1, s_2 \in \tau^{-1}(T')$. Then by definition there exist $t_1, t_2 \in T'$ such that $t_1 \in \tau(s_1)$ and $t_2 \in \tau(s_2)$. Thus $t_1t_2 \in \tau(s_1)\tau(s_2) \subseteq \tau(s_1s_2)$, whence $s_1s_2 \in \tau^{-1}(t_1t_2)$. Therefore $s_1s_2 \in \tau^{-1}(T')$ and hence $\tau^{-1}(T')$ is a subsemigroup of S. □

Example 1.1. Let E be the set of all injective partial functions from $\{1, 2, 3, 4\}$ to itself and let F be the set of all bijections on $\{1, 2, 3, 4\}$. Let τ be the relation that associates to each injective function f the set of all possible bijective extensions of f. For instance, if f is the partial function defined by $f(1) = 3$ and $f(3) = 2$, then $\tau(f) = \{h_1, h_2\}$ were h_1 and h_2 are the bijections given in the following table

	1	**2**	**3**	**4**
h_1	3	1	2	4
h_2	3	4	2	1

Let Id be the identity map on $\{1,2,3,4\}$. Then $\tau^{-1}(\mathrm{Id})$ is the set of *partial identities* on E, listed in the table below:

1	**2**	**3**	**4**
-	-	-	-
-	-	-	4
-	-	3	-
-	-	3	4
-	2	-	-
-	2	-	4
-	2	3	-
-	2	3	4

1	**2**	**3**	**4**
1	-	-	-
1	-	-	4
1	-	3	-
1	-	3	4
1	2	-	-
1	2	-	4
1	2	3	-
1	2	3	4

2 Injective relational morphisms

According to the definition of an injective relation given in Chapter I, a relational morphism $\tau : S \to T$ is *injective* if, for every $s_1, s_2 \in S$, the condition $s_1 \neq s_2$ implies that $\tau(s_1)$ and $\tau(s_2)$ are disjoint, or equivalently, if $\tau(s_1) \cap \tau(s_2) \neq \emptyset$ implies $s_1 = s_2$. Note in particular that if $\alpha : R \to T$ is a surjective morphism, then $\alpha^{-1} : T \to R$ is an injective relational morphism.

Proposition 2.5. *Let $S \xrightarrow{\alpha^{-1}} R \xrightarrow{\beta} T$ be the canonical factorisation of a relational morphism $\tau : S \to T$. Then τ is injective* [*surjective*] *if and only if β is injective* [*surjective*].

Proof. By Proposition I.1.8, α^{-1} is an injective relational morphism. It is also surjective, since $(s,t) \in \alpha^{-1}(s)$ for every $(s,t) \in R$. Thus if β is injective [surjective], then $\tau = \beta \circ \alpha^{-1}$ is also injective [surjective].

Suppose now that τ is injective. Let r_1 and r_2 be two elements of R such that $\beta(r_1) = \beta(r_2) = t$. Since α is surjective, $r_1 \in \alpha^{-1}(\alpha(r_1))$ and $r_2 \in \alpha^{-1}(\alpha(r_2))$. It follows that $t \in \beta\left(\alpha^{-1}\left(\alpha(r_1)\right)\right) \cap \beta\left(\alpha^{-1}\left(\alpha(r_2)\right)\right) = \tau\left(\alpha(r_1)\right) \cap \tau\left(\alpha(r_2)\right)$, whence $\alpha(r_1) = \alpha(r_2)$ since τ is injective. Therefore $r_1 = (\alpha(r_1), \beta(r_1))$ is equal to $r_2 = (\alpha(r_2), \beta(r_2))$.

Finally, if τ is surjective, then β is surjective by Proposition I.1.14. □

Proposition 2.5 has two interesting consequences.

Corollary 2.6. *A semigroup S divides a semigroup T if and only if there exists an injective relational morphism from S to T.*

Proof. If S divides T, there exists a semigroup R, a surjective morphism $\alpha : R \to S$ and an injective morphism $\beta : R \to T$. Then α^{-1} is an injective relational morphism and thus $\tau = \beta \circ \alpha^{-1}$ is an injective relational morphism from S to T.

Conversely, if τ is an injective relational morphism from S to T and if $S \xrightarrow{\alpha^{-1}} R \xrightarrow{\beta} T$ is the canonical factorisation of τ. Proposition 2.5 shows that β is injective. Since α is surjective, S divides T. □

Corollary 2.7. *Let $\tau : S \to T$ be an injective relational morphism. Then for any subsemigroup T' of T, $\tau^{-1}(T')$ divides T'. Furthermore $\tau^{-1}(E(T)) \subseteq E(S)$.*

Proof. Let $S \xrightarrow{\alpha^{-1}} R \xrightarrow{\beta} T$ be the canonical factorisation of τ. Then β is injective by Proposition 2.5 and thus $\beta^{-1}(T')$ is isomorphic to a subsemigroup of T'. Finally, $\tau^{-1}(T')$ is equal to $\alpha(\beta^{-1}(T'))$ and thus divides T'.

Let $s \in \tau^{-1}(E(T))$. Then $\tau(s)$ contains some idempotent f of T. As $\tau(s)\tau(s) \subseteq \tau(s^2)$, $\tau(s^2)$ also contains f. Thus $f \in \tau(s) \cap \tau(s^2)$ whence $s = s^2$ since τ is injective. Thus s is idempotent and $\tau^{-1}(E(T)) \subseteq E(S)$. □

If T is finite, Corollary 2.7 can be improved as follows.

Proposition 2.8. *Let T be a finite semigroup and let $\tau : S \to T$ be an injective relational morphism. Then $\tau^{-1}(E(T)) = E(S)$.*

Proof. Let $e \in E(S)$. By Proposition 1.4, $\tau(e)$ is a subsemigroup of T, which, by Corollary II.6.32, contains an idempotent. Thus $e \in \tau^{-1}(E(T))$, showing that $E(S) \subseteq \tau^{-1}(E(T))$. The opposite inclusion follows from Corollary 2.7. □

3 Relational V-morphisms

Let $\mathbf{V}$ be a variety of finite semigroups. A [relational] morphism $\tau : S \to T$ is said to be a [*relational*] $\mathbf{V}$*-morphism* if, for every subsemigroup T' of T which belongs to $\mathbf{V}$, the semigroup $\tau^{-1}(T')$ also belongs to $\mathbf{V}$.

The definition can be readily adapted to the case of varieties of ordered semigroups. Let $\mathbf{V}$ be a variety of finite ordered semigroups and let S and T be two ordered semigroups. Then a [relational] morphism $\tau : S \to T$ is said to be a [relational] $\mathbf{V}$-morphism if, for every ordered subsemigroup T' of T which belongs to $\mathbf{V}$, the ordered semigroup $\tau^{-1}(T')$ also belongs to $\mathbf{V}$.

In practice, $\mathbf{V}$ is often one of the following varieties:

(1) $\mathbf{A}$, the variety of aperiodic semigroups,

(2) $\mathbf{N}$, the variety of nilpotent semigroups,

(3) $\mathbb{L}\mathbf{1} = [\![ese = e]\!]$, the variety of locally trivial semigroups,

(4) $\mathbb{L}\mathbf{J}^- = [\![e \leqslant ese]\!]$, the variety of ordered semigroups S, such that, for all $e \in E(S)$, the ordered submonoid eSe satisfies the identity $1 \leqslant x$.

A relational $\mathbf{A}$-morphism is also called an *aperiodic relational morphism* and a relational $\mathbb{L}\mathbf{1}$-morphism is also called a *locally trivial relational morphism*.

The definition of a relational $\mathbf{V}$-morphism is formally reminiscent of that of a continuous function. This analogy is confirmed by the following proposition, whose proof is immediate.

Proposition 3.9. *Relational $\mathbf{V}$-morphisms are closed under composition.*

Let us mention another elementary result.

Proposition 3.10. *Injective relational morphisms are relational $\mathbf{V}$-morphisms for every variety $\mathbf{V}$.*

Proof. This follows directly from Corollary 2.7. □

Note that the converse to Proposition 3.10 does not hold. Let $N_2 = \{0, a\}$ and $N_3 = \{0, a, b\}$ be the nilpotent semigroups with two and three elements, respectively and let $\varphi : N_3 \to N_2$ be the morphism defined by $\varphi(a) = \varphi(b) = a$ and $\varphi(0) = 0$. Then the only subsemigroups of N_2 are 0 and N_2. It follows

that φ is a relational $\mathbf{V}$-morphism for every variety $\mathbf{V}$ since $\varphi^{-1}(0) = 0$ and $\varphi^{-1}(N_2) = N_3$, which divides $N_2 \times N_2$. However, φ is not injective.

We can now state our announced result on canonical factorisations.

Proposition 3.11. *Let $S \xrightarrow{\alpha^{-1}} R \xrightarrow{\beta} T$ be the canonical factorisation of a relational morphism $\tau : S \to T$. Then τ is a relational $\mathbf{V}$-morphism if and only if β is a $\mathbf{V}$-morphism.*

Proof. First, α^{-1} is an injective relational morphism and thus a relational $\mathbf{V}$-morphism by Proposition 3.10. Thus if β is a relational $\mathbf{V}$-morphism, then τ is a relational $\mathbf{V}$-morphism by Proposition 3.9.

Conversely, suppose that τ is a relational $\mathbf{V}$-morphism. Let $\gamma : S \times T \to T \times T$ be the relational morphism defined by $\gamma(s,t) = \tau(s) \times \{t\}$. Let T' be a subsemigroup of T belonging to $\mathbf{V}$. Setting $D = \{(t,t) \mid t \in T'\}$, one gets

$$\gamma^{-1}(D) = \{(s,t) \in S \times T \mid t \in \tau(s) \cap T'\} = \beta^{-1}(T')$$

It follows that $\beta^{-1}(T')$ is a subsemigroup of $\tau^{-1}(T') \times T'$ and thus is in $\mathbf{V}$. Thus β is a relational $\mathbf{V}$-morphism. □

Relational morphisms can be restricted to subsemigroups.

Proposition 3.12. *Let $\tau : S \to T$ be a relational morphism and let T' be a subsemigroup of T. Then the relation $\hat{\tau} : \tau^{-1}(T') \to T'$, defined by $\hat{\tau}(s) = \tau(s) \cap T'$, is a relational morphism. Furthermore, if τ is injective [a relational $\mathbf{V}$-morphism], so is $\hat{\tau}$.*

Proof. Let $s \in \tau^{-1}(T')$. Then by definition $\tau(s) \cap T' \neq \emptyset$ and thus $\hat{\tau}(s) \neq \emptyset$. Let $s_1, s_2 \in \tau^{-1}(T')$. One gets

$$\begin{aligned}\hat{\tau}(s_1)\hat{\tau}(s_2) &= (\tau(s_1) \cap T')(\tau(s_2) \cap T') \\ &\subseteq \tau(s_1)\tau(s_2) \cap T' \subseteq \tau(s_1 s_2) \cap T' \subseteq \hat{\tau}(s_1 s_2)\end{aligned}$$

and thus $\hat{\tau}$ is a relational morphism. The second part of the statement is obvious. □

We now turn to more specific properties of relational $\mathbf{V}$-morphisms.

3.1 Aperiodic relational morphisms

Theorem 3.13. *Let $S \xrightarrow{\alpha^{-1}} R \xrightarrow{\beta} T$ be the canonical factorisation of a relational morphism $\tau : S \to T$. The following conditions are equivalent:*

(1) *τ is aperiodic,*
(2) *for every idempotent $e \in T$, $\tau^{-1}(e)$ is aperiodic,*
(3) *the restriction of τ to each group in S is injective,*
(4) *the restriction of τ to each $\mathcal{H}$-class in a regular $\mathcal{D}$-class of S is injective.*

Moreover, one obtains four equivalent conditions (1')–(4') by replacing τ by β and S by R in (1)–(4).

Proof. The equivalence of (1) and (1') follows from Proposition 3.11. Furthermore, (1) implies (2) and (4) implies (3) are obvious.

(3) implies (1). Let T' be an aperiodic subsemigroup of T, $S' = \tau^{-1}(T')$ and let H be a group in S'. Since S, T and R are finite, there exists by Theorem V.5.45 a group H' in R such that $\alpha(H') = H$. Now $\beta(H')$ is a group in T', but since T' is aperiodic, this group is a singleton $\{e\}$. Let $h_1, h_2 \in H$ and $h'_1, h'_2 \in H'$ be such that $\alpha(h_1) = h'_1$ and $\alpha(h'_2) = h_2$. Then $e = \beta(h'_1) = \beta(h'_2) \in \tau(h_1) \cap \tau(h_2)$. It follows from Condition (3) that $h_1 = h_2$, which shows that H is trivial. Therefore S' is aperiodic.

(2) implies (4). Given a regular $\mathcal{H}$-class H, there exists an element $a \in S$ such that the function $h \to ha$ is a bijection from H onto a group G in the same $\mathcal{D}$-class. Let e be the identity of G and let h_1 and h_2 be elements of H such that $\tau(h_1) \cap \tau(h_2) \neq \emptyset$. Then we have

$$\emptyset \neq (\tau(h_1) \cap \tau(h_2))\tau(a) \subseteq \tau(h_1)\tau(a) \cap \tau(h_2)\tau(a) \subseteq \tau(h_1 a) \cap \tau(h_2 a)$$

Setting $g_1 = h_1 a$, $g_2 = h_2 a$ and $g = g_2 g_1^{-1}$, we obtain in the same way

$$\emptyset \neq (\tau(g_1) \cap \tau(g_2))\tau(g_1^{-1}) \subseteq \tau(e) \cap \tau(g)$$

Furthermore, we have

$$\begin{aligned}(\tau(e) \cap \tau(g))(\tau(e) \cap \tau(g)) &\subseteq (\tau(e) \cap \tau(g))\tau(e) \\ &\subseteq \tau(e)\tau(e) \cap \tau(g)\tau(e) \subseteq \tau(ee) \cap \tau(ge) = \tau(e) \cap \tau(g)\end{aligned}$$

which proves that $\tau(e) \cap \tau(g)$ is a nonempty semigroup. Let f be an idempotent of this semigroup. Then $e, g \in \tau^{-1}(f)$, whence $e = g$ since $\tau^{-1}(f)$ is aperiodic. It follows that $g_1 = g_2$ and hence $h_1 = h_2$, which proves (4).

The equivalence of the statements (1)–(4) results from this. Applying this result to β gives the equivalence of (1')–(4'). □

3.2 Locally trivial relational morphisms

Theorem 3.14. *Let $S \xrightarrow{\alpha^{-1}} R \xrightarrow{\beta} T$ be the canonical factorisation of a relational morphism $\tau : S \to T$. The following conditions are equivalent:*

(1) *τ is locally trivial,*

(2) *for every idempotent $e \in T$, $\tau^{-1}(e)$ is locally trivial,*

Moreover, one obtains two equivalent conditions (1')–(2') by replacing τ by β and S by R in (1)–(2).

Proof. The equivalence of (1) and (1') follows from Proposition 3.11. Furthermore, (1) implies (2) is obvious.

(2) implies (1). Up to replacing τ by the relational morphism $\hat{\tau} : S \to \tau(S)$ defined in Proposition 3.12, we may assume that τ is surjective. Furthermore, it follows from Theorem 3.13 that τ is an aperiodic relational morphism.

Let T' be a locally trivial subsemigroup of T and let $S' = \tau^{-1}(T')$. Since T' is an aperiodic semigroup and τ is an aperiodic relational morphism, S' is aperiodic. Let e, f be idempotents of S'. Since $\tau(e)$ and $\tau(f)$ are nonempty subsemigroups of T', there exist idempotents $e', f' \in T'$ such that $e' \in \tau(e)$ and $f' \in \tau(f)$. Now since T' is locally trivial, $e' \mathcal{J} f'$ and thus $e' = a' f' b'$ for some

$a', b' \in T'$. Choose $a, b \in S'$ such that $a' \in \tau^{-1}(a)$ and $b' \in \tau^{-1}(b)$. Then we have

$$e' = a'f'b' \in \tau(a)\tau(f)\tau(b) \subseteq \tau(afb)$$

and therefore $e, afb \in \tau^{-1}(e')$. Since $\tau^{-1}(e')$ is locally trivial by (2), e is in the minimal ideal of $\tau^{-1}(e')$ and hence $e \leqslant_{\mathcal{J}} afb \leqslant_{\mathcal{J}} f$. A dual argument shows that $f \leqslant_{\mathcal{J}} e$ and hence $e \mathcal{J} f$. Thus all the idempotents of S' belong to its minimal ideal and S' is aperiodic. These two properties show that S' is locally trivial.

The equivalence of the statements (1)–(2) results from this. Applying this result to β gives the equivalence of (1')–(2'). □

Proposition 3.15. *Let $\pi : S \to T$ a surjective locally trivial morphism. Then S and T have the same number of regular $\mathcal{J}$-classes.*

Proof. It suffices to show that if x, y are two regular elements of S, $x \mathcal{J} y$ if and only if $\pi(x) \mathcal{J} \pi(y)$. One direction is easy, since π maps a regular $\mathcal{J}$-class onto a regular $\mathcal{J}$-class.

Suppose now that $\pi(x) \mathcal{J} \pi(y)$ and let e and f respectively be idempotents of the $\mathcal{D}$-classes of x and y. Since $e \mathcal{J} x$ and $f \mathcal{J} y$, we also have

$$\pi(e) \mathcal{J} \pi(x) \mathcal{J} \pi(y) \mathcal{J} \pi(f)$$

In particular, $\pi(f) = x\pi(e)y$ for some $x, y \in T$. Since π is surjective, one has $x = \pi(c)$ and $y = \pi(d)$ for some $c, d \in S$. It follows that $\pi(e) = \pi(cfd)$. Now since $\pi(e)$ is idempotent, the semigroup $\pi^{-1}(\pi(e))$ is locally trivial and since e, cfd are both in it, one has $ecfde = e$. Thus $e \leqslant_{\mathcal{J}} f$ and a similar reasoning would show that $f \leqslant_{\mathcal{J}} e$. Therefore $e \mathcal{J} f$, which shows that $x \mathcal{J} y$. □

3.3 Relational $[\![ese \leqslant e]\!]$-morphisms

Recall that if S is an ordered semigroup, the upper set generated by an element $x \in S$ is the set $\uparrow x$ of all $y \in S$ such that $x \leqslant y$.

Proposition 3.16. *Let S be an ordered semigroup and let $e \in E(S)$. Then the ordered semigroup $e(\uparrow e)e$ belongs to the variety $[\![e \leqslant ese]\!]$.*

Proof. Let $R = e(\uparrow e)e$. Let $r \in R$ and $f \in E(R)$. Then $f = ege$ with $e \leqslant g$ and $r = ese$ with $e \leqslant s$. It follows $ef = f = fe$ and $f = fef \leqslant fsf = fesef = frf$. Thus $R \in [\![e \leqslant ese]\!]$. □

Proposition 3.17. *Let $\tau : S \to T$ be a relational morphism. The following conditions are equivalent:*

(1) *τ is a relational $[\![e \leqslant ese]\!]$-morphism,*
(2) *for any $e \in E(T)$, $\tau^{-1}(e(\uparrow e)e)$ is an ordered semigroup of $[\![e \leqslant ese]\!]$,*
(3) *for any $e \in E(T)$, $f \in E(\tau^{-1}(e))$ and $s \in \tau^{-1}(e(\uparrow e)e)$, $f \leqslant fsf$.*

Proof. Proposition 3.16 shows that (1) implies (2) and (2) implies (3) is trivial. Let us show that (3) implies (1). Assuming (3), let R be an ordered subsemigroup of T such that $R \in [\![e \leqslant ese]\!]$. Let $U = \tau^{-1}(R)$, $s \in U$, $r \in \tau(s) \cap R$ and $f \in E(U)$. Since $\tau(f) \cap R$ is a non empty subsemigroup of T, it contains an idempotent e. Now $e \leqslant ere$ since $R \in [\![e \leqslant ese]\!]$ and thus e and ere belong to $e(\uparrow e)e$. Furthermore $f \in \tau^{-1}(e)$, and since $ere \in \tau(f)\tau(s)\tau(f) \subseteq \tau(fsf)$, $fsf \in \tau^{-1}(ere)$. It follows by (3) that $f \leqslant fsf$ and thus $U \in [\![e \leqslant ese]\!]$. Therefore, τ is a relational $[\![e \leqslant ese]\!]$-morphism. □

4 Four examples of V-morphisms

Let M be a monoid and let E be the set of its idempotents. Let 2^E denote the monoid of subsets of E under intersection.

Theorem 4.18. *The following properties hold:*

(1) *If M is $\mathcal{R}$-trivial, the map $\pi : M \to 2^E$ defined by $\pi(s) = \{e \in E \mid es = e\}$ is a $\mathbf{K}$-morphism.*

(2) *If M is $\mathcal{L}$-trivial, the map $\pi : M \to 2^E$ defined by $\pi(s) = \{e \in E \mid se = e\}$ is a $\mathbf{D}$-morphism.*

(3) *If M is $\mathcal{J}$-trivial, the map $\pi : M \to 2^E$ defined by $\pi(s) = \{e \in E \mid es = e = se\}$ is a $\mathbf{N}$-morphism.*

(4) *If M belongs to $\mathbb{D}\mathbf{A}$, the map $\pi : M \to 2^E$ defined by $\pi(s) = \{e \in E \mid ese = e\}$ is a $\mathbb{L}\mathbf{1}$-morphism.*

Proof. (1) If $e \in \pi(s_1 s_2)$, then $es_1 s_2 = e$, whence $e \mathcal{R} es_1$. Since M is $\mathcal{R}$-trivial, $es_1 = e = es_2$ and therefore $e \in \pi(s_1) \cap \pi(s_2)$. In the opposite direction, if $e \in \pi(s_1) \cap \pi(s_2)$, then $es_1 = e = es_2$, whence $es_1 s_2 = e$, that is, $e \in \pi(s_1 s_2)$. Therefore π is a morphism. Fix $A \in 2^E$ and let $x, y \in \pi^{-1}(A)$. Then $x^\omega x = x^\omega$ since M is aperiodic and thus $x^\omega \in \pi(x) = A$. Consequently, $x^\omega y = x^\omega$ since $\pi(y) = A$. It follows that the semigroup $\pi^{-1}(A)$ satisfies the identity $x^\omega y = x^\omega$, and consequently belongs to $\mathbf{K}$. Thus π is a $\mathbf{K}$-morphism.

(2) The proof is similar.

(3) The result follows from (1) and (2) since $\mathbf{N} = \mathbf{K} \cap \mathbf{D}$.

(4) Let $e \in \pi(s_1 s_2)$. Then $es_1 s_2 e = e$, whence $es_1 \mathcal{R} e$ and $s_2 e \mathcal{L} e$ and thus $es_1 e = e = es_2 e$ since the $\mathcal{D}$-class of e only contains idempotents. Conversely if $es_1 e = e = es_2 e$, then $es_1 \mathcal{R} e \mathcal{L} s_2 e$. Since the $\mathcal{D}$-class of e is a semigroup, we have $es_1 s_2 e \in R_{es_1} \cap L_{s_2 e} = R_e \cap L_e = H_e = \{e\}$ and thus $es_1 s_2 e = e$. It follows that π is a morphism. Fix $A \in 2^E$ and let $x, y \in \pi^{-1}(A)$. We conclude as in (1) that $x^\omega y x^\omega = x^\omega$ and therefore $\pi^{-1}(A) \in \mathbb{L}\mathbf{1}$. Thus π is a $\mathbb{L}\mathbf{1}$-morphism. □

5 Mal'cev products

Let $\mathbf{V}$ be a variety of finite semigroups [monoids]. If $\mathbf{W}$ is a variety of semigroups, we define the *Mal'cev product* $\mathbf{W}$ Ⓜ $\mathbf{V}$ to be the class of all finite semigroups [monoids] S such that there exists a $\mathbf{W}$-relational morphism from S to an element of $\mathbf{V}$. If $\mathbf{W}$ is a variety of finite ordered semigroups, we define similarly $\mathbf{W}$ Ⓜ $\mathbf{V}$ to be the class of all finite ordered semigroups [monoids] S such that there exists a $\mathbf{W}$-relational morphism from S to an element of $\mathbf{V}$.

Theorem 5.19. *The following equalities hold:*

$$\mathbf{J} = \mathbf{N} \text{Ⓜ} \mathbf{J_1}, \ \mathbf{R} = \mathbf{K} \text{Ⓜ} \mathbf{J_1}, \ \mathbf{L} = \mathbf{D} \text{Ⓜ} \mathbf{J_1} \text{ and } \mathbb{D}\mathbf{A} = \mathbb{L}\mathbf{1} \text{Ⓜ} \mathbf{J_1}.$$

Proof. The inclusions $\mathbf{J} \subseteq \mathbf{N}$Ⓜ$\mathbf{J_1}$, $\mathbf{R} \subseteq \mathbf{K}$Ⓜ$\mathbf{J_1}$, $\mathbf{L} \subseteq \mathbf{D}$Ⓜ$\mathbf{J_1}$ and $\mathbb{D}\mathbf{A} \subseteq \mathbb{L}\mathbf{1}$Ⓜ$\mathbf{J_1}$ follow from Theorem 4.18. The opposite inclusions can all be proved in the same way. For instance, we give the proof of the inclusion $\mathbb{L}\mathbf{1}$ Ⓜ $\mathbf{J_1} \subseteq \mathbb{D}\mathbf{A}$. If $M \in \mathbb{L}\mathbf{1}$Ⓜ$\mathbf{J_1}$, there exists a locally trivial relational morphism $\tau : M \to N$ with $N \in \mathbf{J_1}$. Let $M \xrightarrow{\alpha^{-1}} R \xrightarrow{\beta} N$ be the canonical factorisation of τ. Then M is a quotient of R and it suffices to verify that $R \in \mathbb{D}\mathbf{A}$. First by Theorem 3.14,

the morphism $\beta : R \to N$ is locally trivial. Let D be a regular $\mathcal{D}$-class of R. Then $\beta(D)$ is a regular $\mathcal{D}$-class of N. Since N is idempotent and commutative, $\beta(D)$ is reduced to an idempotent e and therefore $D \subseteq \tau^{-1}(e)$. Since $\tau^{-1}(e)$ is locally trivial, D only contains idempotents. Therefore $R \in \mathbb{D}\mathbf{A}$. □

TO DO: other examples.

6 Three examples of relational morphisms

In this section, we give three examples of relational morphisms stemming from the theory of automata and recognisable languages. Our first example describes an important property of the concatenation product. The second one deals with purity, a property of the star of a language. The third one gives a nice syntactic property of flower automata.

6.1 Concatenation product

For $0 \leqslant i \leqslant n$, let L_i be a recognizable language of A^*, let $\eta_i : A^* \to M(L_i)$ be its syntactic morphism and let

$$\eta : A^* \to M(L_0) \times M(L_1) \times \cdots \times M(L_n)$$

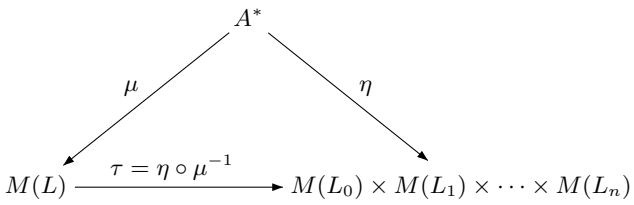
be the morphism defined by

$$\eta(u) = (\eta_0(u), \eta_1(u), \ldots, \eta_n(u))$$

Let a_1, a_2, …, a_n be letters of A and let $L = L_0a_1L_1 \cdots a_nL_n$. Let $\mu : A^* \to M(L)$ be the syntactic morphism of L. The properties of the relational morphism

$$\tau = \eta \circ \mu^{-1} : M(L) \to M(L_0) \times M(L_1) \times \cdots \times M(L_n)$$

were first studied by Straubing [157] and later in [99, 118, 108].

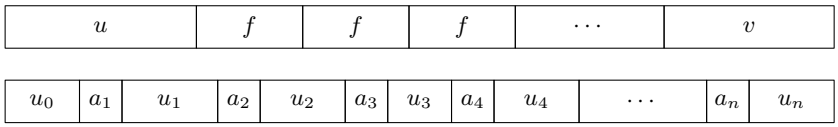


Theorem 6.20. *The relational morphism $\tau : M(L) \to M(L_0) \times M(L_1) \times \cdots \times M(L_n)$ is a relational $[\![e \leqslant ese]\!]$-morphism.*

Proof. Let R be an ordered subsemigroup of $M(L_0) \times M(L_1) \times \cdots \times M(L_n)$ satisfying the identity $x^\omega \leqslant x^\omega y x^\omega$, and let $x, y \in \eta^{-1}(R)$. Let k be an integer such that $\mu(x^k)$ and $\eta(x^k)$ are idempotent. It suffices to show that for every $u, v \in A^*$, $ux^kv \in L$ implies $ux^kyx^kv \in L$. Let $r = 2n+1$. Then $\eta(x^{rk}) = \eta(x^k)$, and since $ux^kv \in L$, $ux^{rk}v \in L$. Consequently, there is a factorisation of the form $ux^{rk}v = u_0a_1 \cdots a_nu_n$, where $u_i \in L_i$ for $0 \leqslant i \leqslant n$. The next step is a lemma of independent interest.

Lemma 6.21. *Suppose that $uf^rv = u_0a_1\cdots a_nu_n$, with $r > n$. Then one of the words u_i contains f as a factor.*

Proof. Otherwise each factor f should contain at least one letter a_i.



Since $r > n$, this would give a contradiction. □

By Lemma 6.21 applied to $f = x^k$, there exist $1 \leqslant h \leqslant n$ and $0 \leqslant j \leqslant r-1$ such that $u_h = u'_hx^{2k}u''_h$ for some $u'_h, u''_h \in A^*$, $ux^{jk} = u_0a_1\cdots a_{h-1}u'_h$ and $x^{(r-j-1)k}v = u''_ha_h\cdots a_nu_n$. Now since $\eta(x)$ and $\eta(y)$ belong to R,

$$\eta(x^k) \leqslant \eta(x^k)\eta(y)\eta(x^k)$$

and by projection onto $M(L_h)$, $\eta_h(x^k) \leqslant \eta_h(x^k)\eta_h(y)\eta_h(x^k)$. In particular, the condition $u'_hx^ku''_h \in L_h$ implies $u'_hx^kyx^ku''_h \in L_h$. Thus $ux^{(j+1)k}yx^{(r-j)k}v \in L$ and hence $ux^kyx^kv \in L$, which concludes the proof. □

Theorem 6.20 is often used in the following weaker form.

Corollary 6.22. *The relational morphism $\tau : M(L) \to M(L_0) \times M(L_1) \times \cdots \times M(L_n)$ is an aperiodic relational morphism.*

Proof. By Theorem 6.20, τ is a relational $[\![e \leqslant ese]\!]$-morphism. In particular, for each idempotent e, $\tau^{-1}(e)$ is a semigroup satisfying the identity $e \leqslant ese$. In particular, it satisfies the identity $x^\omega \leqslant x^\omega xx^\omega$, that is $x^\omega \leqslant x^{\omega+1}$ and is aperiodic by Proposition VI.2.2. Thus τ is aperiodic. □

Let $L_0, L_1, \ldots, L_n$ be languages of A^* and let $a_1, \ldots, a_n$ be letters of A. The (marked) product

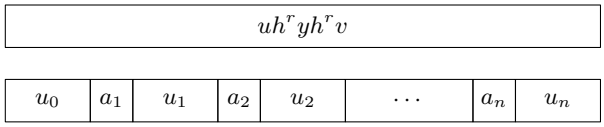
$$L = L_0a_1L_1\cdots a_nL_n$$

is said to be *unambiguous* if every word of L admits a unique decomposition of the form $u = u_0a_1u_1\cdots a_nu_n$ with $u_0 \in L_0$, …, $u_n \in L_n$.

Example 6.1. Let $A = \{a, b, c\}$. The marked product $\{a, c\}^*a\{1\}b\{b, c\}^*$ is unambiguous.

Theorem 6.23. *If the product $L_0a_1L_1\cdots a_nL_n$ is unambiguous, the relational morphism $\tau : M(L) \to M(L_0)\times M(L_1)\times\cdots\times M(L_n)$ is a locally trivial relational morphism.*

Proof. By Theorem 3.14, it suffices to show that if e is an idempotent of $M(L_0)\times M(L_1) \times \cdots \times M(L_n)$, then the semigroup $\tau^{-1}(e)$ is locally trivial. It follows from Theorem 6.20 that $\tau^{-1}(e)$ satisfies the identity $x^\omega \leqslant x^\omega yx^\omega$ and it just remains to prove the opposite identity $x^\omega yx^\omega \leqslant x^\omega$. Let $x, y \in \eta^{-1}(e)$, let k be an integer such that $\mu(x^k)$ is idempotent and let $h = x^k$. It suffices to show that $uhyhv \in L$ implies $uhv \in L$. Let $r = n + 2$. Then $\eta(h^r) = \eta(h)$, and if $uhyhv \in L$, then $uh^ryh^rv \in L$. Consequently, there is a factorisation of the form $uh^ryh^rv = u_0a_1\cdots a_nu_n$, where $u_i \in L_i$ for $0 \leqslant i \leqslant n$.



First assume that one of the words u_i contains hyh as a factor, that is, $u_i = u_i' hyh u_i''$ with $u_0 a_1 \cdots a_i u_i' = uh^{r-1}$ and $u_i'' a_{i+1} \cdots a_n u_n = h^{r-1} v$. Since $\eta(x) = \eta(y) = e$, one has $\eta_i(hyh) = \eta_i(h)$ and hence, $u_i' hyh u_i'' \in L_i$ implies $u_i' h u_i'' \in L_i$. Consequently, one has

$$uh^{2r-1}v = uh^{r-1}hh^{r-1}v = u_0 a_1 \cdots a_i (u_i' h u_i'') a_{i+1} \cdots a_n u_n$$

which shows that $uh^{2r-1}v$ belongs to L. Since $\eta(h^{2r-1}) = \eta(h)$, it follows that uhv is also in L, as required.

Suppose now that none of the words u_i contains hyh as a factor. Then there are factorisations of the form

$$uh^{r-1} = u_0 a_1 \cdots a_i u_i' \qquad hyh = u_i'' a_{i+1} \cdots a_j u_j' \qquad h^{r-1}v = u_j'' a_{j+1} \cdots a_n u_n$$

with $u_i = u_i' u_i''$ and $u_j = u_j' u_j''$. By Lemma 6.21 applied to the word $u_0 a_1 \cdots a_i u_i'$, one of the words $u_0, \ldots, u_{i-1}, u_i'$ contains h as a factor. Similarly, one of the words $u_j'', u_{j+1}, \ldots, u_n$ contains h as a factor. Therefore one gets factorisations

$$\begin{aligned} uh^{r-1} &= u_0 a_1 \cdots a_\ell u_\ell' h u_\ell'' a_{\ell+1} \cdots u_i' \\ h^{r-1}v &= u_j'' a_{j+1} \cdots a_m u_m' h u_m'' a_{m+1} \cdots a_n u_n \end{aligned}$$

with

$$u_0 a_1 \cdots a_\ell u_\ell' = uh^p \qquad u_\ell' h u_\ell'' = u_\ell \qquad u_\ell'' \cdots u_i' = h^{r-p-1}$$

and

$$u_j'' a_{j+1} \cdots a_m = h^q \qquad u_m' h u_m'' = u_m \qquad u_m'' \cdots a_n u_n = h^{r-q-1} v$$

Since $h \sim_{L_\ell} h^{r-p} y h^p$ and $u_\ell' h u_\ell'' \in L_\ell$, one also has $u_\ell' h^{r-p} y h^p u_\ell'' \in L_\ell$. By a similar argument, one gets $u_m' h^{r-q-1} y h^{q+1} u_m'' \in L_m$. Finally, the word $uh^r yh^r yh^r v$ can be factored either as

$$u_0 a_1 \cdots u_{\ell-1} a_\ell (u_\ell' h^{r-p} y h^p u_\ell'') a_{\ell+1} \cdots a_n u_n$$

or as

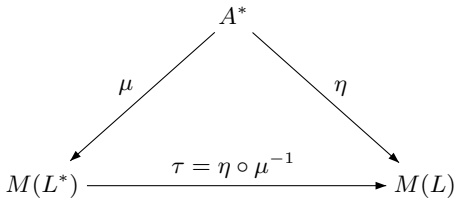
$$u_0 a_1 \cdots a_m (u_m' h^{r-q-1} y h^{q+1} u_m'') a_{m+1} \cdots a_n u_n$$

a contradiction, since this product should be unambiguous. □

6.2 Pure languages

A submonoid M of A^* is *pure* if for all $u \in A^*$ and $n > 0$, the condition $u^n \in M$ implies $u \in M$.

Let $\eta : A^* \to M(L)$ be the syntactic morphism of L and $\mu : A^* \to M(L^*)$ be the syntactic morphism of L^*. Then $\tau = \eta \circ \mu^{-1}$ is a relational morphism from $M(L^*)$ to $M(L)$.



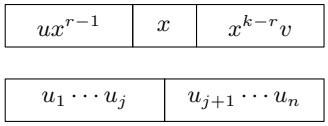
The following result is due to Straubing [157].

Theorem 6.24. *If L^* is pure, then the relational morphism $\tau : M(L^*) \to M(L)$ is aperiodic.*

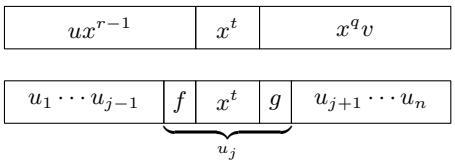
Proof. Let e be an idempotent of $M(L)$ and let $x \in \eta^{-1}(e)$. Let k be an integer such that $k > |x|$ and $\mu(x^k)$ is idempotent. By Proposition VI.2.2, it suffices to show that for every $u, v \in A^*$,

$$ux^k v \in L^* \text{ implies } ux^{k+1}v \in L^* \tag{6.1}$$

Suppose that $ux^k v \in L^*$. Then $ux^k v = u_1 \cdots u_n$, where each u_i belongs to $L - \{1\}$. Let us say that the r-th occurrence of x is *cut* if, for some j, ux^{r-1} is a prefix of $u_1 \cdots u_j$ and $u_1 \cdots u_j$ is a proper prefix of ux^r.



There are two cases to consider. First assume that one of the occurrences of x is not cut. Then for some $j \in \{1, \ldots, n\}$, the word u_j contains x as a factor, that is, $ux^{r-1} = u_1 \cdots u_{j-1} f$, $u_j = fx^t g$ and $x^q v = g u_{j+1} \cdots u_n$ for some $f, g \in A^*$ and $t > 0$ such that $r + t + q - 1 = k$.



Since $x \sim_L x^2$ and since $t > 0$, one gets $fx^t g \sim_L fx^{t+1}g$ and thus $fx^{t+1}g \in L$. It follows that $ux^{k+1}v = u_1 \cdots u_{j-1} fx^{t+1} g u_{j+1} \cdots u_n \in L^*$, proving (6.1) in this case.

Suppose now that every occurrence of x is cut. Then for $1 \leqslant r \leqslant k$, there exists $j_r \in \{1, \ldots, n\}$ and $f_r \in A^*$, $g_r \in A^+$ such that

$$ux^{r-1} f_r = u_1 \cdots u_j, \; x = f_r g_r \text{ and } g_r x^{k-r} v = u_{j_{r+1}} \cdots u_n$$

Since there are $|x|$ factorisations of x of the form fg, and since $|x| < k$, there exist two indices $r \neq r'$ such that $f_r = f_{r'}$ and $g_r = g_{r'}$. Thus, for some indices $i < j$ and some factorisation $x = fg$, one has $ux^{r-1}f = u_1 \cdots u_i$, $gx^s f = u_{i+1} \cdots u_j$ and $gx^t v = u_{j+1} \cdots u_n$. It follows that $gx^s f = g(fg)^s f = (gf)^{s+1}$. Since $gx^s f \in L^*$ and since L^* is pure, $gf \in L^*$. Therefore, $ux^{k+1}v = ux^{r-1}xx^s xxx^t v = (u^{r-1}f)(gx^s f)(gf)(gx^t v) \in L^*$, proving (6.1) in this case as well. □

Corollary 6.25. *If L is star-free and L^* is pure, then L^* is star-free.*

Proof. By Theorem VI.3.3, L is star-free if and only if $M(L)$ is aperiodic. Now, if L^* is pure, the relational morphism τ is aperiodic and hence $M(L^*)$ is aperiodic. It follows that L^* is star-free. □

6.3 Flower automata

Let L be a finite language of A^+. The *flower automaton* of L^+ is the finite nondeterministic automaton $\mathcal{A} = (Q, A, E, I, F)$, where $Q = \{1, 1\} \cup \{(u, v) \in A^+ \times A^+ \mid uv \in L\}$, $I = F = \{(1, 1)\}$. There are four types of transitions:

$$\begin{aligned}
&\{((u, av) \xrightarrow{a} (ua, v)) \mid uav \in L,\ (u, v) \neq (1, 1)\}\\
&\{((u, a) \xrightarrow{a} (1, 1)) \mid ua \in L,\ u \neq 1\}\\
&\{((1, 1) \xrightarrow{a} (a, v)) \mid av \in L,\ v \neq 1\}\\
&\{((1, 1) \xrightarrow{a} (1, 1)) \mid a \in L\}
\end{aligned}$$

It is easy to see that this automaton recognises L^+.

Example 6.2. Let $A = \{a, b\}$ and $L = \{a, ba, aab, aba\}$.

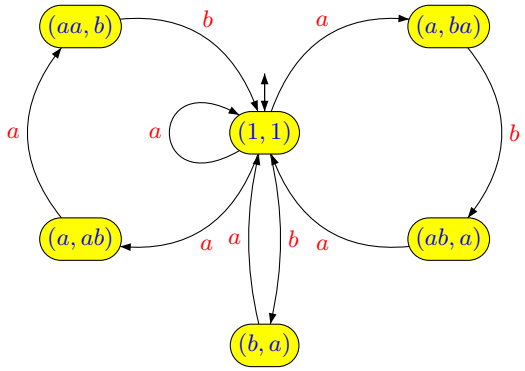
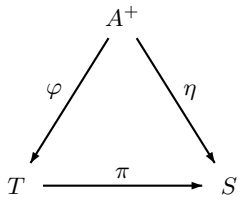


Figure 6.1. A flower automaton.

The transition semigroup T of the flower automaton of L^+ is called the *flower semigroup* of L^+. Let S be the syntactic semigroup of L^+. Since T recognises L^+, there is a surjective morphism π from T onto S.

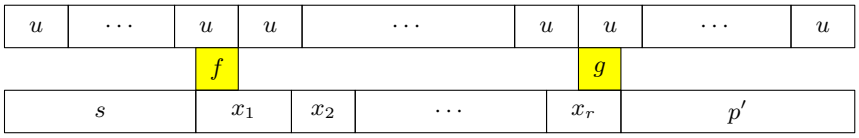
Theorem 6.26. *Let L be a finite language. The natural morphism from the flower semigroup of L^+ onto its syntactic semigroup is a locally trivial morphism.*

Proof. Let $\mathcal{A}$ be the flower automaton of L^+, let T be its transition semigroup and let $\varphi : A^+ \to T$ be the natural morphism from A^+ onto T. Let also $\eta = \pi \circ \varphi$ be the syntactic morphism of L^+.



Let e be an idempotent of S and let $f, s \in \pi^{-1}(e)$, with f idempotent. By Theorem 3.14, it suffices to show that $fsf = f$. Let u and v be nonempty words such that $\varphi(u) = f$ and $\varphi(v) = s$. Then $\pi(u) = \pi(v) = e$ and thus $v \sim_{L^+} u \sim_{L^+} u^2$. Now, the condition $fsf = f$ is equivalent to stating that uvu and u have the same behaviour in $\mathcal{A}$. Furthermore, since $\varphi(u)$ is idempotent, the words u^n, for $n > 0$, have the same behaviour as u in $\mathcal{A}$. Thus it is sufficient to prove that the words $u^n vu$ and u^n have the same behaviour in $\mathcal{A}$, for some sufficiently large n (the precise value of n will be determined later in the proof). Let (p, s) and (p', s') be two states of $\mathcal{A}$ and let N be the maximal length of the words of L. Then any path from (p, s) to (p', s') of length $\geqslant 3N$ has to visit the state $(1, 1)$ at least twice. In particular, if w is a word of length $\geqslant 3N$, there is a path $(p, s) \xrightarrow{w} (p', s')$ if and only if $w \in sL^+p'$.

Taking $n = 3N$, it suffices now to prove that the conditions $u^n \in sL^+p'$ and $u^n vu \in sL^+p'$ are equivalent. Suppose that $u^n \in sL^+p'$. Then $u^n = sx_1x_2 \cdots x_rp'$ for some $x_1, x_2, \ldots, x_r \in L$. Let i be the smallest integer such that s is a prefix of u^i, let j be the smallest integer such that p' is a suffix of u^j and let f and g be the words defined by the conditions $sf = u^i$ and $gp' = u^j$.



Setting $k = n - i - j$ we get $fu^kg = x_1 \cdots x_r$ and thus $fu^kg \in L^+$. Now since $v \sim_{L^+} u \sim_{L^+} u^2$, we also have $fu^{n-i}vu^{n-j}g \in L^+$ and thus $sfu^{n-i}vu^{n-j}gp' = u^iu^{n-i}vu^{n-j}u^j = u^nvu^n$. Therefore $u^nvu^n \in sL^+p'$.

Conversely, if $u^nvu^n \in sL^+p'$, then $u^nvu^n = (sf)u^{n-i}vu^{n-j}(gp')$, with $fu^{n-i}vu^{n-j}g \in L^+$. Again, since $v \sim_{L^+} u \sim_{L^+} u^2$, we get $fu^kg \in L^+$ and finally $u^n \in sL^+p'$. □

Chapter XVII

Unambiguous star-free languages

Recall that $\mathbb{D}\mathbf{A}$ denotes the class of finite semigroups in which every regular $\mathcal{D}$-class is an aperiodic semigroup (or idempotent semigroup, which is equivalent in this case). Several characterisations of $\mathbb{D}\mathbf{A}$ were given in Proposition XI.4.29.

1 Unambiguous star-free languages

Let A be a finite alphabet. The set of *unambiguous star-free* languages of A^* is the smallest set of languages of A^* containing the languages of the form B^*, for $B \subseteq A$, which is closed under finite union and unambiguous marked product.

Let us start by an elementary observation.

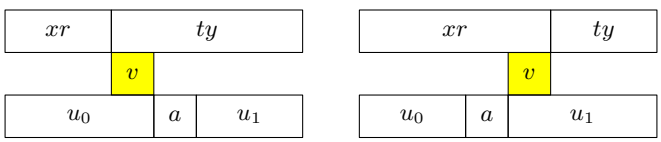
Proposition 1.1. *Every finite language is unambiguous star-free.*

Proof. If $a_1, \ldots, a_k$ are letters of A, the marked product $\{1\}a_1\{1\}a_2 \cdots a_k\{1\}$ is unambiguous. It follows that for any word u, the language $\{u\}$ is unambiguous star-free. Furthermore, any finite language is the disjoint union of the languages $\{u\}$, for $u \in F$. Thus every finite language is unambiguous star-free. □

Example 1.1. The language $\{a, c\}^* a \{1\} b \{b, c\}^*$ is unambiguous star-free (see Example XVI.6.1).

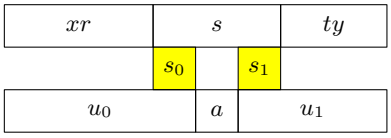
Proposition 1.2. *Let $L = L_0 a L_1$ be an unambiguous marked product of two languages of A^* and let $r, s, t \in A^*$. If $r \sim_{L_0} rs$ and $st \sim_{L_1} t$, then $rst \sim_L rt$.*

Proof. Let $x, y \in A^*$ and suppose that $xrty \in L$. Then there exists exactly one pair $(u_0, u_1) \in L_0 \times L_1$ such that $xrty = u_0 a u_1$. Thus either xr is a prefix of u_0 or ty is a suffix of u_1.



In the first case, one gets $xrv = u_0$ and $vau_1 = ty$ for some $v \in A^*$ and since $r \sim_{L_0} rs$ and $xrv \in L_0$, it follows that $xrsv \in L_0$ and thus $xrsty = xrsvau_1 \in L$. The second case is similar.

Suppose now that $xrsty \in L$. Then there exists exactly one pair $(u_0, u_1) \in L_0 \times L_1$ such that $xrsty = u_0au_1$. If xrs is a prefix of u_0 or if sty is a suffix of u_1, one can use the same argument as above to show that $xrty \in L$. The critical case occurs when $s = s_0as_1$, $xrs_0 = u_0$ and $s_1ty = u_1$.



Then $xrs_0 \in L_0$ and hence $xrss_0 \in L_0$ since $r \sim_{L_0} rs$. Similarly, $s_1sty \in L_1$ since $st \sim_{L_1} t$. It follows that $(xrss_0)a(s_1ty) = xrssty = (xrs_0)a(s_1sty)$. This contradicts the fact that the marked product L_0aL_1 is unambiguous. □

The aim of this section is to prove the following theorem.

Theorem 1.3. *A language is unambiguous star-free if and only if its syntactic monoid is finite and belongs to* $\mathbb{D}\mathbf{A}$.

Proof. The easiest part of the proof relies on Theorem XVI.6.23. Let

$$L = L_0a_1L_1 \cdots a_nL_n$$

be an unambiguous marked product. Let $M_0, \ldots, M_n$ and M be the respective syntactic monoids of $L_0, \ldots, L_n$ and L.

Lemma 1.4. *If* $M_0, \ldots, M_n$ *belong to* $\mathbb{D}\mathbf{A}$, *so does* M.

Proof. Since the monoids $M_0, \ldots, M_n$ are all in $\mathbb{D}\mathbf{A}$, so is their product N. Furthermore, by Theorem XVI.6.23, there is a locally trivial relational morphism τ from M to N. Let $M \xrightarrow{\alpha^{-1}} R \xrightarrow{\beta} N$ be the canonical factorisation of τ. Since M is a quotient of R, it suffices to prove that R belongs to $\mathbb{D}\mathbf{A}$. By Theorem XVI.3.14, β is a locally trivial morphism. Let D be a regular $\mathcal{D}$-class of R. Then $\beta(D)$ is a regular $\mathcal{D}$-class of N and since N belongs to $\mathbb{D}\mathbf{A}$, this $\mathcal{D}$-class is an aperiodic simple semigroup. In particular, $\beta(D)$ is in $\mathbb{L}\mathbf{1}$ and since β is a locally trivial morphism, the semigroup $\beta^{-1}(\beta(D))$ also belongs to $\mathbb{L}\mathbf{1}$. It follows in particular that D is an aperiodic simple semigroup and thus R belongs to $\mathbb{D}\mathbf{A}$. □

Chapter XVIII

Wreath product

In this chapter, we introduce two important notions: the semidirect product of semigroups and the wreath product of transformation semigroups. We also prove some basic decomposition results.

1 Semidirect product

Let S and T be semigroups. We write the product in S additively to provide a more transparent notation, but it is not meant to suggest that S is commutative. A *left action* of T on S is a map $(t, s) \mapsto t \cdot s$ from $T^1 \times S$ to S such that, for all $s, s_1, s_2 \in S$ and $t, t_1, t_2 \in T$,

(1) $t_1 \cdot (t_2 \cdot s) = (t_1 t_2) \cdot s$

(2) $t \cdot (s_1 + s_2) = t \cdot s_1 + t \cdot s_2$

(3) $1 \cdot s = s$

If S is a monoid with identity 0, the action is *unitary* if it satisfies, for all $t \in T$,

(4) $t \cdot 0 = 0$

The *semidirect product* of S and T (with respect to the given action) is the semigroup $S * T$ defined on $S \times T$ by the multiplication

$$(s, t)(s', t') = (s + t \cdot s', tt')$$

2 Wreath product

Let $X = (P, S)$ and $Y = (Q, T)$ be two transformation semigroups. To make the notation more readable, we shall denote the semigroup S and its action on P additively and the semigroup T and its action on Q multiplicatively. The *wreath product* of X and Y, denoted by $X \circ Y$, is the transformation semigroup $(P \times Q, W)$ where W consists of all pairs (f, t), with f is a function from Q to S and $t \in T$. Since we are thinking of f as acting on the right on Q, we will use the more suitable notation $q \cdot f$ in place of $f(q)$. The action of W on $P \times Q$ is given by

$$(p, q) \cdot (f, t) = (p + q \cdot f, q \cdot t) \tag{2.1}$$

We claim that this action is faithful. Indeed, if $(p, q) \cdot (f, t) = (p, q) \cdot (f', t')$ for all $(p, q) \in P \times Q$, then $q \cdot t = q \cdot t'$ for all $q \in Q$ and thus $t = t'$ since T acts faithfully

on Q. On the other hand, $p+q\cdot f = p+q\cdot f'$ for all $p \in P$ and thus $q\cdot f = q\cdot f'$ since S acts faithfully on P. Thus $f = f'$, proving the claim. In particular W can be considered as a subset of the semigroup of all transformations on $P \times Q$. We leave it to the reader to verify that W is closed under composition and that the product on W is defined by

$$(f,t)(f',t') = (g,tt')$$

where g is defined, for each $q \in Q$ by

$$q\cdot g = q\cdot f + (q\cdot t)\cdot f'$$

Let us now verify that Formula (2.1) really defines an action of W on $P \times Q$. If $(p,q) \in P \times Q$ and $(f,t),(f',t') \in W$, we have

$$\begin{aligned}\big((p,q)\cdot(f,t)\big)\cdot(f',t') = (p+q\cdot f, q\cdot t)\cdot(f',t') &= (p+q\cdot f+(q\cdot t)\cdot f', q\cdot tt')\\ &= (p,q)\big((f,t)(f',t')\big)\end{aligned}$$

Given two semigroups S and T, consider the wreath product $(S^1,S)\circ(T^1,T) = (S^1\times T^1, W)$. The semigroup W is called the *wreath product* of S and T and is denoted by $S\circ T$. The connections with the semidirect product and the product are given in the next propositions.

Proposition 2.1. *Let S and T be semigroups. Then every semidirect product of S and T is a subsemigroup of $S\circ T$. Furthermore, $S\circ T$ is a semidirect product of S^{T^1} and T.*

Proof. Let $S * T$ be a semidirect product of S and T. Let $\varphi\colon S*T \to S\circ T$ be the function defined by $\varphi(s,t) = (f,t)$ where $f\colon T^1 \to S$ is given by $t\cdot f = t\cdot s$ for every $t \in T^1$. It is easy to verify that φ is a semigroup morphism.

For the second part of the statement, define a left action $(t,f) \mapsto t\cdot f$ of T on S^{T^1} as follows: $t\cdot f$ is the function from T^1 to S defined by $t'\cdot(t\cdot f) = (t't)\cdot f$. Then the semidirect product defined by this action is isomorphic to $S\circ T$. □

Proposition 2.2. *Let X and Y be transformation semigroups. Then $X\times Y$ divides $X\circ Y$.*

Proof. Let $X = (P,S)$ and $Y = (Q,T)$. Since the transformation semigroups $X\times Y$ and $X\circ Y$ have the same set of states, $P\times Q$, it suffices to show that $S\times T$ can be embedded to $S^Q \times T$. With each pair (s,t), associate the pair (f,t), where f is the constant map onto s. Then, for every pair $(p,q) \in P\times Q$, $(p,q)\cdot(s,t) = (p+s, q\cdot t) = (p+q\cdot f, q\cdot t) = (p,q)\cdot(f,t)$, which concludes the proof. □

A routine computation shows that the wreath product on transformation semigroups is associative. The wreath product also preserves division.

Proposition 2.3. *If (P_1,S_1) divides (Q_1,T_1) and (P_2,S_2) divides (Q_2,T_2), then $(P_1,S_1)\circ(P_2,S_2)$ divides $(Q_1,T_1)\circ(Q_2,T_2)$.*

Proof. Let $\pi_1 : Q_1 \to P_1$ and $\pi_2 : Q_2 \to P_2$ be the surjective mappings defining the divisions. Let $\pi = \pi_1 \times \pi_2 : Q_1 \times Q_2 \to P_1 \times P_2$. For $(f, s_2) \in (P_1, S_1) \circ (P_2, S_2)$, define $\widehat{(f, s_2)} = (g, \hat{s}_2)$ by choosing a cover $\hat{s}_2$ of s_2 and, for each $q_2 \in Q_2$, a cover $g(q_2)$ of $f(\pi_2(q_2))$. Now, for each $(q_1, q_2) \in Q_1 \times Q_2$,

$$\begin{aligned}\pi(q_1, q_2) \cdot (f, s_2) &= (\pi_1(q_1), \pi_2(q_2)) \cdot (f, s_2) = (\pi_1(q_1) \cdot f(\pi_2(q_2)), \pi_2(q_2) \cdot s_2) \\ &= (\pi_1(q_1 \cdot g(q_2)), \pi_2(q_2 \cdot \hat{s}_2)) = \pi(q_1 \cdot g(q_2), q_2 \cdot \hat{s}_2) \\ &= \pi((q_1, q_2) \cdot (g, \hat{s}_2))\end{aligned}$$

and this computation concludes the proof. □

In view of Proposition 2.3, we have the following corollary.

Corollary 2.4. *If S_1 divides T_1 and S_2 divides T_2, then $S_1 \circ S_2$ divides $T_1 \circ T_2$.*

If $X = (P, S)$ is a transformation semigroup, then X^1 denotes the transformation semigroup obtained by adjoining to S the identity map 1_P on P. If p is a state, we let c_p denote the *constant map* defined, for all $q \in P$, by $c_p(q) = p$. The transformation semigroup obtained by adjoining to S all the constant maps c_p is denoted by $\overline{X}$.

Proposition 2.5. *Let X and Y be transformation semigroups. Then $(X \circ Y)^1$ divides $X^1 \circ Y^1$ and $\overline{X \circ Y}$ divides $\overline{X} \circ \overline{Y}$.*

Proof. Let $X = (P, S)$ and $Y = (Q, T)$. First note that the four transformation semigroups $\overline{X \circ Y}$, $\overline{X} \circ \overline{Y}$, $(X \circ Y)^1$ and $X^1 \circ Y^1$ have the same set of states, $P \times Q$. Next, $1_{P \times Q}$ has the same action as $(f, 1_Q) \in (S^1)^Q \times T$, where $f(q) = 1_P$ for all $q \in Q$. Thus $(X \circ Y)^1$ embeds to $X^1 \circ Y^1$.

Finally, if $(p, q) \in P \times Q$, the constant map $c_{(p,q)}$ has exactly the same action as the pair $(g, c_q) \in \overline{S}^Q \times \overline{T}$ where $g(x) = c_p$ for all $x \in Q$. Thus $\overline{X \circ Y}$ embeds into $\overline{X} \circ \overline{Y}$. □

3 Basic decomposition results

In this section, we give some useful decomposition results. Let us first remind Proposition XI.4.21, which gives a useful decomposition result for commutative monoids. Useful decompositions involving $\tilde{U}_n$ and U_n are given in the next propositions.

Proposition 3.6. *For every $n > 0$, U_n divides U_2^{n-1} and $\tilde{U}_n$ divides $\tilde{U}_2^n$.*

Proof. Arguing by induction on n, it suffices to verify that U_n divides $U_{n-1} \times U_2$. Actually a simple computation shows that U_n is isomorphic to the submonoid N of $U_{n-1} \times U_2$ defined as follows:

$$N = \{(1, 1)\} \cup \{(a_i, a_1) \mid 1 \leqslant i \leqslant n - 1\} \cup \{(a_1, a_2)\}$$

A dual proof works for $\tilde{U}_n$. □

A more precise result follows from Proposition XIV.1.6: a monoid is idempotent and $\mathcal{R}$-trivial if and only if it divides $\tilde{U}_2^n$ for some $n > 0$. Dually, a monoid is idempotent and $\mathcal{L}$-trivial if and only if it divides U_2^n for some $n > 0$.

Proposition 3.7. *For every $n > 0$, $\tilde{U}_n$ divides $U_n \circ U_2$.*

Proof. Let $\pi : U_n \times U_2 \to \tilde{U}_n$ be the surjective partial map defined by $\pi(1, a_1) = 1$ and, for $1 \leqslant i \leqslant n$, $\pi(a_i, a_2) = a_i$.

For $1 \leqslant j \leqslant n$, we set $\hat{a}_j = (f_j, a_2)$ where $f_j : U_2 \to U_n$ is defined by $1 \cdot f_j = a_2 \cdot f_j = 1$ and $a_1 \cdot f_j = a_j$. We also set $\hat{1} = (f, 1)$ where $f : U_2 \to U_n$ is defined by $1 \cdot f = a_1 \cdot f = a_2 \cdot f = 1$. Now a simple verification shows that π is indeed a cover:

$$\begin{aligned}
\pi(a_i, a_2) \cdot 1 &= \pi(a_i, a_2) = \pi(a_i + a_2 \cdot f, a_2 \cdot 1) = \pi((a_i, a_2) \cdot \hat{1}) \\
\pi(1, a_1) \cdot 1 &= \pi(1, a_1) = \pi(1 + a_1 \cdot f, a_1 \cdot 1) = \pi((1, a_1)(f, 1)) = \pi((1, a_1) \cdot \hat{1}) \\
\pi(a_i, a_2) \cdot a_j &= a_i = \pi(a_i, a_2) = \pi(a_i + a_2 \cdot f_j, a_2 \cdot a_2) = \pi((a_i, a_2) \cdot \hat{a}_j) \\
\pi(1, a_1) \cdot a_j &= a_j = \pi(a_j, a_2) = \pi(1 + a_1 \cdot f_j, a_1 \cdot a_2) = \pi((1, a_1) \cdot \hat{a}_j)
\end{aligned}$$

Thus $\tilde{U}_n$ divides $U_n \circ U_2$. □

The following result now follows immediately from Propositions 3.6 and 3.7.

Corollary 3.8. *For every $n > 0$, $\tilde{U}_n$ divides $\underbrace{U_2 \circ \cdots \circ U_2}_{n+1 \text{ times}}$.*

For each $n > 0$, let $\mathbf{D_n}$ be the class of finite semigroups S such that, for all s_0, s_1, …, s_n in S, $s_0 s_1 \cdots s_n = s_1 \cdots s_n$. In such a semigroup, a product of more than n elements is determined by the last n elements. By Proposition XI.4.15, these semigroups are lefty trivial. We shall now give a decomposition result for the semigroups in $\mathbf{D_n}$. As a first step, we decompose $\bar{\mathbf{n}}$ as a product of copies of $\bar{\mathbf{2}}$.

Lemma 3.9. *If $2^k > n$, then $\bar{\mathbf{n}}$ divides $\bar{\mathbf{2}}^k$.*

Proof. The result is trivial, since if T is any subset of size n of $\bar{\mathbf{2}}^k$, (T, T) is a sub-transformation semigroup of $\bar{\mathbf{2}}^k$ isomorphic to $\bar{\mathbf{n}}$. □

We now decompose the semigroups of $\mathbf{D_n}$ as an iterated wreath product of transformation semigroups of the form (T, T).

Proposition 3.10. *Let S be a semigroup of $\mathbf{D_n}$ and let $T = S \cup \{t\}$, where t is a new element. Then (S^1, S) divides $\underbrace{(T, T) \circ \cdots \circ (T, T)}_{n \text{ times}}$.*

Proof. Let $\varphi : T^n \to S^1$ be the partial function defined on sequences of the form $(t, \ldots, t, x_i, \ldots, x_1)$, where $x_1, \ldots, x_i \in S$, by

$$\varphi(t, \ldots, t, x_i, \ldots, x_1) = \begin{cases} x_i \cdots x_1 & \text{if } i > 0 \\ 1 & \text{if } i = 0 \end{cases}$$

Clearly φ is surjective. If $s \in S$, we set $\hat{s} = (f_{n-1}, \ldots, f_1, s)$, where, for $1 \leqslant i \leqslant n-1$, $f_i : T^i \to T$ is defined by $(t_i, \ldots, t_1) \cdot f_i = t_i$. Thus

$$(t_n, \ldots, t_1)\hat{s} = (t_{n-1}, \ldots, t_1, s)$$

It follows that if $p = (t, \ldots, t, x_i, \ldots, x_1)$ is in the domain of φ, then $p \cdot \hat{s}$ is also in the domain of φ and $\varphi(p \cdot \hat{s}) = \varphi(p) \cdot s$. This proves the proposition. □

Proposition 3.10 and Lemma 3.9 now give immediately.

Corollary 3.11. *Every semigroup of* $\mathbf{D_n}$ *divides a wreath product of copies of* $\overline{\mathbf{2}}$.

$\mathcal{R}$-trivial monoids admit also a simple decomposition.

Theorem 3.12. *A monoid is $\mathcal{R}$-trivial if and only if it divides a wreath product of the form $U_1 \circ \cdots \circ U_1$.*

Proof. We first show that every monoid of the form $U_1 \circ \cdots \circ U_1$ is $\mathcal{R}$-trivial. Since U_1 itself is $\mathcal{R}$-trivial, and since, by Proposition 2.1, a wreath product is a special case of a semidirect product, it suffices to show that the semidirect product $S * T$ of two $\mathcal{R}$-trivial monoids S and T is again $\mathcal{R}$-trivial. Indeed, consider two $\mathcal{R}$ equivalent elements (s,t) and (s',t') of $S * T$. Then, $(s,t)(x,y) = (s',t')$ and $(s',t')(x',y') = (s,t)$ for some elements (x,y) and (x',y') of $S * T$. Therefore, on one hand $s + tx = s'$ and $s' + t'x' = s$ and on the other hand, $ty = t'$ and $t'y' = t$. It follows that $s \mathcal{R} s'$ and $t \mathcal{R} t'$. Therefore $s = s'$ and $t = t'$, and $S * T$ is $\mathcal{R}$-trivial.

Let $M = \{s_1, \ldots, s_n\}$ be an $\mathcal{R}$-trivial monoid of size n. We may assume that $s_i \leqslant_{\mathcal{R}} s_j$ implies $j \leqslant i$. Let us identify the elements of U_1^n with words of length n on the alphabet $\{0,1\}$. Let $\varphi : U_1 \times \cdots \times U_1 \to M$ be the onto partial function defined by

$$\varphi(1^{n-j}0^j) = s_j \quad (0 \leqslant j \leqslant n)$$

Thus $\varphi(u)$ is not defined if $u \notin 1^*0^*$. For each $s \in M$, let

$$\hat{s} = (f_{n-1}, \ldots, f_2, a_1)$$

where

$$a_1 = \begin{cases} 1 & \text{if } s = 1 \\ 0 & \text{if } s \neq 1 \end{cases}$$

and $f_{i+1} : \underbrace{U_1 \times \cdots \times U_1}_{i \text{ times}} \to U_1$ is defined by

$$f_{i+1}(1^{i-j}0^j) = \begin{cases} 1 & \text{if } s_j s = s_k \text{ and } k \leqslant i \\ 0 & \text{if } s_j s = s_k \text{ and } k > i \end{cases}$$

If $u \notin 1^*0^*$, the value of $f_{i+1}(u)$ can be chosen arbitrarily.

Let $p = 1^{n-j}0^j$ and $s \in M$. Let k be such that $s_k = s_j s$. Since $s_k \leqslant_{\mathcal{R}} s_j$, $k \geqslant j$. Then

$$\begin{aligned} p\hat{s} &= (f_{n-1}, \ldots, f_2, a_1)(1^{n-j}0^j) \\ &= 1^{n-k}0^k \end{aligned}$$

whence $\varphi(p\hat{s}) = s_k = s_j s = \varphi(p)s$. Therefore, M divides $U_1 \circ \cdots \circ U_1$. □

As a preparation to the next theorem, we prove another decomposition result, which is important in its own right.

□

Proposition 3.13. *Let M be a finite aperiodic monoid and let $S = M - \{1\}$. Then at least one of the following cases occurs:*

(1) *M is a monogenic monoid,*
(2) *M is isomorphic to $\tilde{U}_n$ for some $n > 0$,*
(3) *there is a proper left ideal I of S and a proper subsemigroup T of S such that $S = I \cup T$.*

Proof. Since M is aperiodic, S is a subsemigroup of M. If S is empty, the result is trivial. Otherwise, let $L_1, L_2, \ldots, L_n$ be the maximal $\mathcal{L}$-classes of S for the $\mathcal{L}$-order on $\mathcal{L}$-classes.

If $n > 1$, then $S - L_n$ is a proper left ideal of S. Indeed, let $s \in S - L_n$ and $t \in M$. Then $ts \in S - L_n$, since otherwise $ts \in L_n$, $ts \leqslant_{\mathcal{L}} s$ and $s \in L_n$ by the maximality of L_n. Let T be the subsemigroup generated by L_n. Then T is a proper subsemigroup since $T \cap L_1 = \emptyset$. Moreover $S = (S - L_n) \cup T$ by construction.

We are left with the case $n = 1$: S has a unique maximal $\mathcal{L}$-class L. Let T be the subsemigroup generated by L. Then $S = (S - L) \cup T$ and $S - L$ is a left ideal of S different from S. If $S - L$ is empty, then L is the minimal ideal of S and we are in case (2), otherwise $S - L$ is a proper left ideal of S. We now inspect the case $T = S$. If L is a singleton, then S is monogenic and we are in case (1). If $|L| > 1$, then by Proposition V.1.18, L is regular and consists of $\mathcal{L}$-equivalent idempotents. It follows that $T = L = S$ and we are again in case (2). □

Proposition 3.14. *Let M be a monoid. Suppose that $M = L \cup N$ where L is a left ideal and N is a submonoid of M. Then M divides $L^1 \circ \overline{N}$.*

Proof. Let $\varphi : L^1 \times N \to M$ be the map defined by $\varphi(l, n) = ln$. Since $M = L \cup N$ and $L \cup N \subseteq L^1 N$, $M = L^1 N$ and φ is onto.

Recall that $\overline{N} = N \cup \{c_n \mid n \in N\}$, where one extends the product on N by setting $sc_t = c_s c_t = c_t$ and $c_s t = c_{st}$ for all $s, t \in N$.

Let us construct a covering of M by $L^1 \circ \overline{N}$. Let $m \in M$ and let f and g be the functions from $\overline{N}$ to L^1 defined, for all $n \in N$, by

$$\begin{aligned} f(n) &= 1 & \qquad f(c_n) &= 1 \\ g(n) &= nm & \qquad g(c_n) &= 1 \end{aligned}$$

We now define $\widehat{m} \in L^1 \circ \overline{N}$ by setting

$$\widehat{m} = \begin{cases} (g, c_1) & \text{if } m \in L \\ (f, m) & \text{if } m \notin L \end{cases}$$

Let $(l, n) \in L^1 \times N$. Then

$$\varphi(l, n) \cdot m = (ln) \cdot m = lnm$$

Let $m \in M$. If $m \in L$, we get

$$(l, n) \cdot \widehat{m} = (l, n)(g, c_1) = (l \cdot g(n), n \cdot c_1) = (l \cdot g(n), 1) = (lnm, 1)$$

and since L is a left ideal, $lnm \in L$. On the other hand, if $m \in N$,

$$(l, n) \cdot \widehat{m} = (l, n)(f, m) = (l \cdot f(n), n \cdot m) = (l, nm)$$

and since N is a monoid, $nm \in N$. In both cases, $\varphi((l, n) \cdot \widehat{m}) = lnm$. It follows that φ is a covering and thus M divides $L^1 \circ \overline{N}$. □

Theorem 3.15. *A monoid is aperiodic if and only if it divides a wreath product of the form* $U_2 \circ \cdots \circ U_2$.

Proof. Let M be an aperiodic monoid and let $S = M - \{1\}$. Consider the three cases given by Proposition 3.13. If M is monogenic, then it is $\mathcal{R}$-trivial, and the result follows from Theorem 3.12. If M is isomorphic to $\tilde{U}_n$ for some $n > 0$, the result follows from Corollary 3.8. Finally, suppose there is a proper left ideal L of S and a proper subsemigroup T of S such that $S = L \cup T$. Then L is also a left ideal of M and $N = T \cup \{1\}$ is a proper submonoid of M. Thus by Proposition 3.14, M divides $L^1 \circ \overline{N}$.

Arguing by induction on $|M|$, we may assume that L and N divide wreath products of copies of U_2. It follows, by Proposition 2.5, that L^1 alsos divide wreath products of copies of U_2, since $U_2 = U_2^1$ and that $\overline{N}$ divides wreath products of copies of $\overline{U}_2 = U_3$. Now, by Proposition 3.6, U_3 divides $U_2 \circ U_2$. Finally, M itself divides a wreath product of copies of U_2. □

Proposition 3.16. *Let* $X = (P, S)$ *be a transformation semigroup such that* $P \cdot S = P$. *Then* $\bar{\mathbf{2}} \circ X$ *divides* $X \circ (R, R)$, *where* R *is the set* $\{1, 2\}^P \times S$.

Proof. Define $\varphi : P \times R \to \{1, 2\} \times P$ by setting $\varphi(p, f, s) = (p \cdot f, p \cdot s)$ for each $p \in P$, $f \in \{1, 2\}^P$ and $s \in S$. Given a transformation $v = (g, t)$ of $\bar{\mathbf{2}} \circ X$, with $g \in \{1, 2\}^P$ and $t \in S$, define the transformation $\hat{v}$ of $X \circ (R, R)$ by setting

$$(p, f, s) \cdot \hat{v} = (p \cdot s, g, t)$$

then we have

$$\begin{aligned} \varphi(p, f, s) \cdot v = (p \cdot f, p \cdot s)(g, t) &= (p \cdot f + (p \cdot s) \cdot g, p \cdot st) \\ ((p \cdot s) \cdot g, p \cdot st) &= \varphi(p \cdot s, g, t) = \varphi((p, f, s) \cdot \hat{v}) \end{aligned}$$

Thus $\bar{\mathbf{2}} \circ X$ divides $X \circ (R, R)$. □

Given a variety $\mathbf{V}$, we say that a semigroup S is *locally* in a variety $\mathbf{V}$ if the local semigroup of each idempotent is in $\mathbf{V}$. For instance, a semigroup S is *locally trivial* if, for each $s \in S$ and $e \in E(S)$, $ese = e$.

We shall present without proof our last decomposition result (see the Notes section).

Proposition 3.17. *A semigroup is locally* $\mathcal{R}$*-trivial if and only if it divides a wreath product of the form* $U_1 \circ \cdots \circ U_1 \circ \bar{\mathbf{2}} \circ \cdots \circ \bar{\mathbf{2}}$.

Proof. TO DO. □

We now turn to groups.

Proposition 3.18. *Let* $\pi : G \to H$ *be a surjective morphism of groups and let* $K = \pi^{-1}(1)$. *Then* G *is isomorphic to a subgroup of* $K \circ H$.

Proof. For each $h \in H$, select an element p_h of G such that $\pi(p_h) = h$. For each $g \in G$, define a map $\hat{g} : H \to K$ by setting $\hat{g}(h) = p_h g p_{h\pi(g)}^{-1}$. Finally, let $\varphi : G \to K \circ H$ be the map defined by

$$\varphi(g) = (\hat{g}, \pi(g))$$

Let us show that φ is a group morphism. By Proposition II.3.6, it suffices to prove that φ is a semigroup morphism. Let $g_1, g_2 \in G$ and let $s_1 = \pi(g_1)$ and $s_2 = \pi(g_2)$. One gets

$$\varphi(g_1)\varphi(g_2) = (\hat{g}_1, s_1)(\hat{g}_2, s_2) = (\hat{g}_1 + s_1\hat{g}_2, s_1s_2)$$

with, for each $h \in H$,

$$(\hat{g}_1 + s_1\hat{g}_2)(h) = \hat{g}_1(h)\hat{g}_2(hs_1) = p_h g_1 p_{hs_1}^{-1} p_{hs_1} g_2 p_{hs_1s_2}^{-1} = p_h g_1 g_2 p_{hs_1s_2}^{-1} = \widehat{g_1g_2}(h)$$

and thus $\varphi(g_1)\varphi(g_2) = \varphi(g_1g_2)$.

Finally, we prove that φ is injective. If $\varphi(g_1) = \varphi(g_2)$, then $\pi(g_1) = \pi(g_2) = s$ and $\hat{g}_1(1) = \hat{g}_2(1)$, that is, $p_1g_1p_s^{-1} = p_1g_2p_s^{-1}$, whence $g_1 = g_2$. It follows that G is isomorphic to a subgroup of $K \circ H$. □

A subgroup H of a group G is *normal* if, for each h in H and each g in G, the element hgh^{-1} is still in H. A group G is *soluble* if there is an ascending chain of subgroups $1 = G_0 \subseteq G_1 \subseteq \cdots \subseteq G_n = G$ such that, for $1 \leqslant i \leqslant n$, G_{i-1} is normal in G_i and G_i/G_{i-1} is a commutative group.

Proposition 3.19. *A group is soluble if and only if it divides a wreath product of cyclic groups.*

Proof. TO DO. □

Theorem 3.20 (Krohn-Rhodes)**.** *Let M be a monoid. Then there exists a sequence $G_1, \ldots, G_n$ of groups dividing M and a sequence $M_0, \ldots, M_n$ of aperiodic monoids such that M divides $M_0 \circ G_1 \circ M_1 \cdots \circ G_n \circ M_n$.*

Proof. TO DO. □

4 Exercises

Section 3

Exercise 1. Show that any finite inverse monoid divides a semidirect product of the form $S * G$, where S an idempotent and commutative monoid and G is a finite group. Actually, a stronger result holds: a finite monoid divides the semidirect product of an idempotent and commutative monoid by a group if and only if its idempotents commute.

Chapter XIX

Sequential functions

So far, we have only used automata to define languages but more powerful models allow one to define functions or even relations between words. These automata not only read an input word but they also produce an output. Such devices are also called transducers. In the deterministic case, which forms the topic of this chapter, they define the so called sequential functions. The composition of two sequential functions is also sequential.

Straubing's "wreath product principle" [155, 160] provides a description of the languages recognised by the wreath product of two monoids. It has numerous applications, including Schützenberger's theorem on star-free languages [33, 88], the characterisation of languages recognised by solvable groups [155] or the expressive power of fragments of temporal logic [32, 176].

1 Definitions

1.1 Pure sequential transducers

A *pure sequential transducer* is an 6-tuple $\mathcal{T} = (Q, A, R, q_0, \cdot, *)$, where Q is a finite set of *states*, A is a finite alphabet called the *input alphabet*, R is a semigroup (possibly infinite) called the *output semigroup*, $q_0 \in Q$ is the *initial state*, $(q, a) \mapsto q \cdot a \in Q$ and $(q, a) \mapsto q * a \in R$ are partial functions with the same domain contained in $Q \times A$, called respectively the *transition function* and the *output function*. Both functions are conveniently represented in Figure 1.1, in which the vertical slash is a separator.

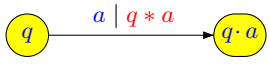


Figure 1.1. A transition and its output.

The transition and the output functions can be extended to partial functions $Q \times A^* \to Q$ (resp. $Q \times A^* \to R^1$) by setting, for each $u \in A^*$ and each $a \in A$:

$$q \cdot 1 = q \qquad\qquad q * 1 = 1$$

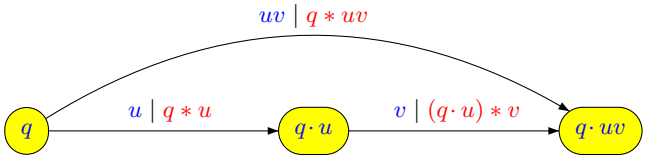
$$\begin{aligned} q\cdot(ua) &= (q\cdot u)\cdot a && \text{if } q\cdot u \text{ and } (q\cdot u)\cdot a \text{ are defined} \\ q*(ua) &= (q*u)((q\cdot u)*a) && \text{if } q*u,\ q\cdot u \text{ and } (q\cdot u)*a \text{ are defined} \end{aligned}$$

To make this type of formulas more readable, it is convenient to fix some precedence rules on the operators. Our choice is to give highest priority to concatenation, then to dot and then to star. For instance, we write $q\cdot ua$ for $q\cdot(ua)$, $q*ua$ for $q*(ua)$ and $q\cdot u*a$ for $(q\cdot u)*a$.

Proposition 1.1. *Let $\mathcal{T} = (Q, A, R, q_0, \cdot, *)$ be a pure sequential transducer. Then the following formulas hold for all $q \in Q$ and for all $u, v \in A^*$:*

(1) $q\cdot uv = (q\cdot u)\cdot v$

(2) $q*uv = (q*u)(q\cdot u*v)$



Proof. The first formula is a standard property of deterministic automata. The second one can be proved by induction on the length of v. If $v = 1$, it is obvious. If $v = wa$, where a is a letter, one gets successively

$$\begin{aligned} q*uv = q*uwa &= (q*uw)(q\cdot uw*a) \\ &= (q*u)(q\cdot u*w)((q\cdot u)\cdot w*a) \\ &= (q*u)(q\cdot u*wa) = (q*u)(q\cdot u*v) \end{aligned}$$

which gives the result. □

The function *realised* by the pure sequential transducer $\mathcal{T}$ is the partial function $\varphi\colon A^* \to R^1$ defined by

$$\varphi(u) = q_0 * u$$

A *pure sequential function* is a partial function that can be realised by a pure sequential transducer. Pure sequential functions preserve prefixes: if u is a prefix of v, and if $\varphi(v)$ is defined, then $\varphi(u)$ is defined and is a prefix of $\varphi(v)$.

Example 1.1. The transducer pictured in Figure 1.2 converts text from uppercase to lowercase.

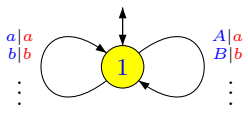


Figure 1.2. Conversion from upper to lower case letters.

Example 1.2. The pure sequential automaton of Figure 1.3 removes unnecessary spaces from text. More precisely, it replaces any series of spaces with a

single space. In the automaton, space is represented by an underline and the dot represents any character other than space.

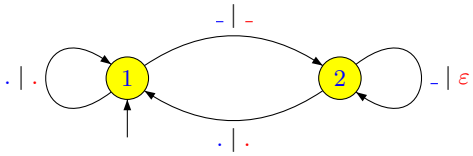


Figure 1.3. Removing repeated spaces.

Example 1.3. Consider the prefix code $P = \{0000, 0001, 001, 010, 011, 10, 11\}$ represented by the tree pictured in Figure 1.4.

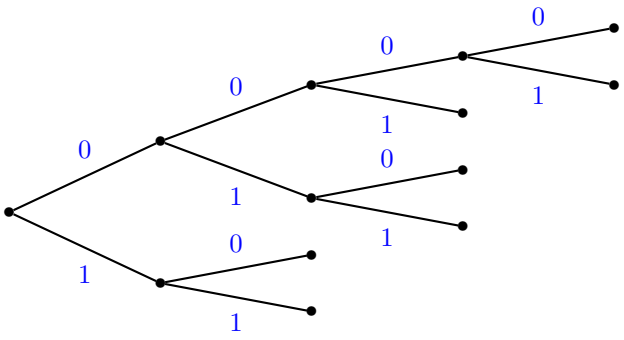


Figure 1.4. A prefix code.

Since P is a prefix code, the coding function $\varphi : \{a, b, c, d, e, f\}^* \to \{0, 1\}^*$ defined by

$$\begin{array}{llll} \varphi(a) = 0000 & \varphi(b) = 0001 & \varphi(c) = 001 & \varphi(d) = 010 \\ \varphi(e) = 011 & \varphi(f) = 10 & \varphi(g) = 11 & \end{array}$$

is injective. It thus performs a *coding* of the alphabet $\{a, b, c, d, e, f\}$ by the alphabet $\{0, 1\}$. *Decoding* can then be carried out using the pure sequential automaton shown in Figure 1.5 which can then be represented in the more traditional form of Figure 1.6.

This example is generic. If one starts from a coding function of carried out using a finite prefix code, the decoding function is a pure sequential function.

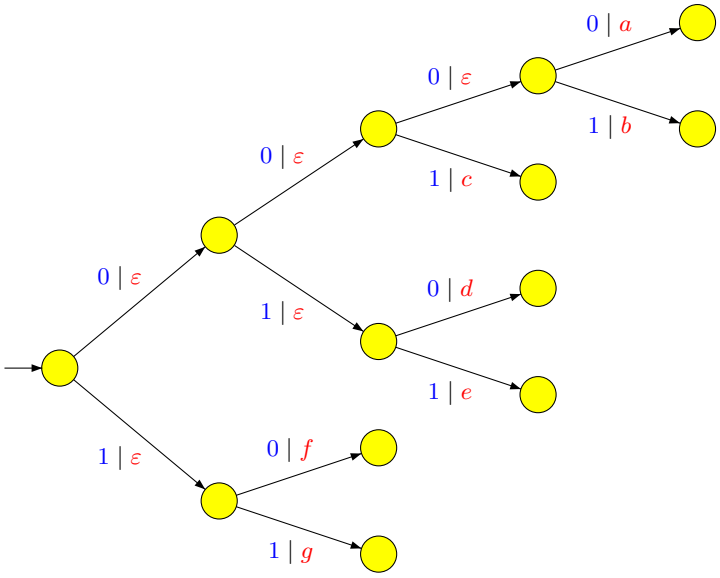


Figure 1.5. A pure sequential automaton realizing the decoding.

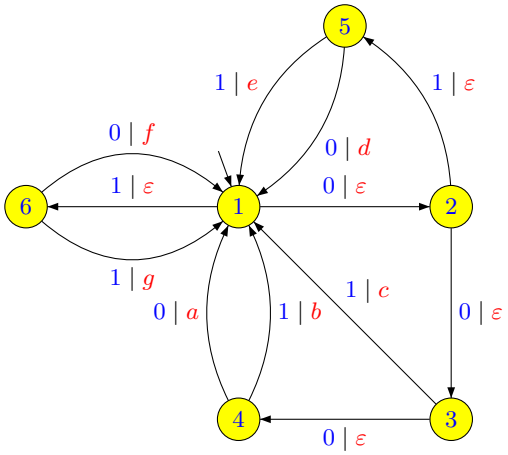


Figure 1.6. The same pure sequential automaton...

Example 1.4. Let $\varphi : \{a, b\}^* \to (\mathbb{N}, +)$ be the function which counts the number of occurrences of the factor aba in a word. It is a pure sequential function, realised by the transducer represented in Figure 1.7.

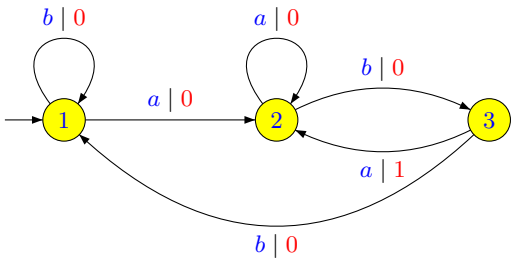


Figure 1.7. A function computing the number of occurrences of *aba*.

1.2 Sequential transducers

Let a and b be two distinct letters. A function as simple as the function $u \to ua$ is not pure sequential since b is a prefix of bb, but ba is not a prefix of bba. However, it is easy to realise this function by a machine which reproduces its input as its output and concatenates an a at the end of the final output. Such a machine is an example of a sequential transducer. Here is the formal definition.

A *sequential transducer* is an 8-tuple $\mathcal{T} = (Q, A, R, q_0, \cdot, *, m, \rho)$, where $(Q, A, R, q_0, \cdot, *)$ is a pure sequential transducer, $m \in R^1$ is the *initial prefix* and $\rho : Q \to R^1$ is a partial function, called the *terminal function*.

The function *realised* by the sequential transducer $\mathcal{T}$ is the partial function $\varphi : A^* \to R^1$ defined by

$$\varphi(u) = m(q_0 * u)\rho(q_0 \cdot u)$$

A *sequential function* is a partial function that can be realised by a sequential transducer.

Example 1.5. Let $u, v \in A^*$. The map $\varphi : A^* \to A^*$ defined by $\varphi(x) = uxv$ is realised by the sequential transducer represented in Figure 1.8. Its initial prefix is u and its terminal function ρ is defined by $\rho(1) = v$.

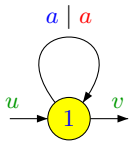


Figure 1.8. A sequential automaton realizing the function $x \to uxv$.

Example 1.6. The function $\varphi : A^* \to A^*$ defined by $\varphi(x) = x(ab)^{-1}$ is realised by the sequential transducer represented in Figure 1.9.

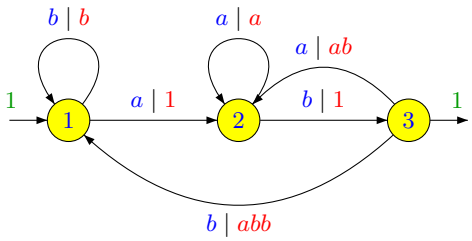


Figure 1.9. A sequential automaton realizing the function $x \to x(ab)^{-1}$.

Example 1.7. In the *reverse binary representation*, a binary word $a_0a_1\cdots a_n$ represents the number $\sum_{0\leqslant i\leqslant n} a_i 2^i$. For instance, the word 1011 represents $1+4+8=13$. Multiplication by 3 can then be performed by the following sequential transducer.

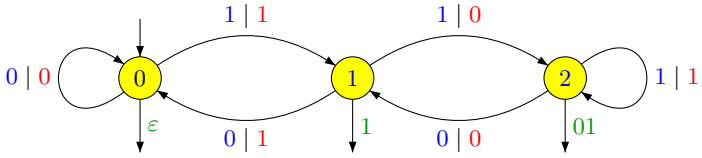


Figure 1.10. A sequential transducer realising the multiplication by 3.

For instance, on the input 1011, which represents 13, the output would be 111001, which represents $1+2+4+32=39$.

2 Composition of sequential functions

The goal of this section is to establish that the composition of two [pure] sequential functions is also a [pure] sequential function.

Let $\varphi : A^* \to B^*$ and $\theta : B^* \to C^*$ be two pure sequential functions, realised respectively by the pure sequential transducers

$$\mathcal{A} = (Q, A, B, q_0, \cdot, *) \text{ and } \mathcal{B} = (P, B, C, p_0, \cdot, *)$$

The composition of these two functions can be realised by feeding $\mathcal{B}$ with the output of $\mathcal{A}$. This observation leads to the following definition. The *wreath product* of $\mathcal{B}$ by $\mathcal{A}$ is the pure sequential transducer

$$\mathcal{B} \circ \mathcal{A} = (P \times Q, A, C, (p_0, q_0), \cdot, *)$$

defined by

$$(p,q)\cdot a = (p\cdot(q * a), q\cdot a) \text{ and } (p,q) * a = p * (q * a)$$

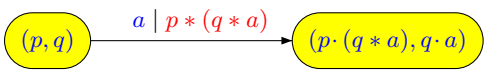


Figure 2.1. A transition and its output in $\mathcal{B} \circ \mathcal{A}$.

Intuitively, the second component q of the pair (p, q) allows one to simulate the transitions of the automaton $\mathcal{A}$. To obtain the first component and to compute the output, we start from a transition of $\mathcal{A}$,

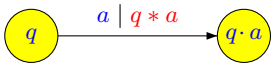


Figure 2.2. A transition and its output in $\mathcal{A}$.

and then we take the output and use it as input of $\mathcal{B}$.

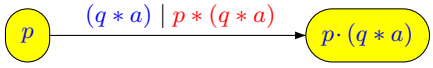


Figure 2.3. The output of $\mathcal{A}$ as input of $\mathcal{B}$.

This definition can be extended to sequential transducers as follows. Let

$$\mathcal{A} = (Q, A, B, q_0, \cdot, *, n, \rho) \text{ and } \mathcal{B} = (P, B, C, p_0, \cdot, *, m, \sigma)$$

be two sequential transducers. The *wreath product* of $\mathcal{B}$ by $\mathcal{A}$ is the sequential transducer

$$\mathcal{B} \circ \mathcal{A} = (P \times Q, A, C, (p_0 \cdot n, q_0), \cdot, *, m(p_0 * n), \omega)$$

defined by

$$\begin{aligned}(p, q) \cdot a &= (p \cdot (q * a), q \cdot a)\\ (p, q) * a &= p * (q * a)\\ \omega(p, q) &= (p * \rho(q))\sigma(p \cdot \rho(q))\end{aligned}$$

The calculation of the initial prefix and the terminal function are shown in Figure 2.4.

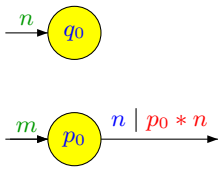


Figure 2.4. The initial prefix.

The next proposition describes the behaviour of $\mathcal{B} \circ \mathcal{A}$.

Proposition 2.2. *The following formulas hold for all states (p, q) of $\mathcal{B} \circ \mathcal{A}$ and for all $u \in A^*$:*

$$\begin{aligned}(p, q) \cdot u &= (p \cdot (q * u), q \cdot u)\\ (p, q) * u &= p * (q * u)\end{aligned}$$

Proof. We prove the result by induction on the length of u. If $u = 1$, one gets on the one hand $(p, q)\cdot 1 = (p, q)$ and on the other hand $q * 1 = 1$ and $q\cdot 1 = q$, whence $(p\cdot(q * 1), q\cdot 1) = (p\cdot 1, q) = (p, q)$, which proves the first formula. The second one is trivial since by definition $(p, q) * 1 = 1 = p * (q * 1)$.

Suppose by induction that the formulas are true for a word u of length n and let a be a letter. Setting $v = ua$, $w = q * u$ and $w' = (q\cdot u) * a$, one gets $q * v = q * (ua) = (q * u)((q\cdot u) * a) = ww'$ and

$$\begin{aligned}(p, q)\cdot v &= (p\cdot(q * u), q\cdot u)\cdot a = (p\cdot w, q\cdot u)\cdot a\\ &= (p\cdot w\cdot(q\cdot u * a), (q\cdot u)\cdot a) = (p\cdot ww', q\cdot v) = (p\cdot(q * v), q\cdot v)\end{aligned}$$

which proves the first formula. Furthermore

$$\begin{aligned}(p, q) * v &= \big((p, q) * u\big)\big((p, q)\cdot u * a\big) = \big(p * (q * u)\big)\Big(\big(p\cdot(q * u), q\cdot u\big) * a\Big)\\ &= (p * w)\big((p\cdot w, q\cdot u) * a\big) = (p * w)\big((p\cdot w) * (q\cdot u * a)\big)\\ &= (p * w)\big((p\cdot w) * w'\big) = p * (ww') = p * (q * v)\end{aligned}$$

which gives the second one. □

Example 2.1. Let $A = \{a, b\}$, $B = \{a, b\}$ and $C = \{a, b, c\}$ and let $\mathcal{A}$ and $\mathcal{B}$ be the pure sequential transducers represented in Figure 2.5.

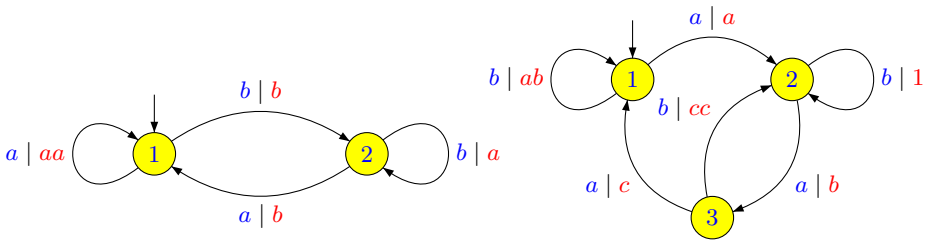


Figure 2.5. The automata $\mathcal{A}$ (on the left) and $\mathcal{B}$ (on the right).

The wreath product $\mathcal{B}\circ\mathcal{A}$ is defined by the formula

$(1,1)\cdot a = (3,1)$	$(1,1) * a = ab$	$(1,1)\cdot b = (1,2)$	$(1,1) * b = ab$
$(1,2)\cdot a = (1,1)$	$(1,2) * a = ab$	$(1,2)\cdot b = (2,2)$	$(1,2) * b = a$
$(2,1)\cdot a = (1,1)$	$(2,1) * a = bc$	$(2,1)\cdot b = (2,2)$	$(2,1) * b = 1$
$(2,2)\cdot a = (2,1)$	$(2,2) * a = 1$	$(2,2)\cdot b = (3,2)$	$(2,2) * b = b$
$(3,1)\cdot a = (2,1)$	$(3,1) * a = ca$	$(3,1)\cdot b = (2,2)$	$(3,1) * b = cc$
$(3,2)\cdot a = (2,1)$	$(3,2) * a = cc$	$(3,2)\cdot b = (1,2)$	$(3,2) * b = c$

and is represented in Figure 2.6.

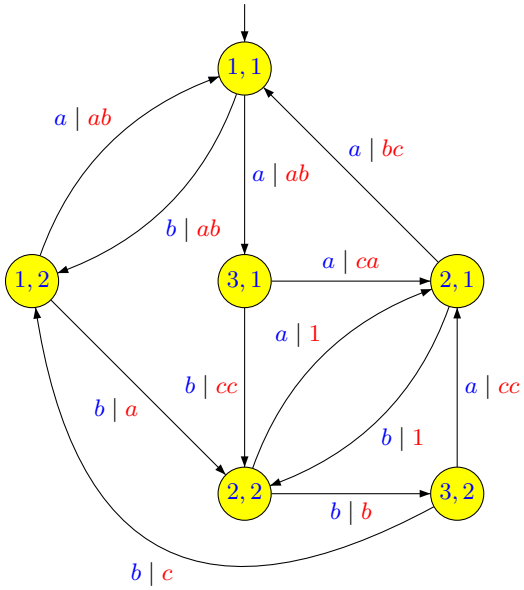


Figure 2.6. The wreath product $\mathcal{B} \circ \mathcal{A}$.

Theorem 2.3. *Let $\mathcal{A}$ and $\mathcal{B}$ be two [pure] sequential transducers realising the functions $\varphi : A^* \to B^*$ and $\theta : B^* \to C^*$. Then $\mathcal{B} \circ \mathcal{A}$ realises the function $\theta \circ \varphi$.*

Proof. Let $\mathcal{A} = (Q, A, B, q_0, \cdot, *, n, \rho)$ and $\mathcal{B} = (P, B, C, p_0, \cdot, *, m, \sigma)$ be two sequential transducers realising the fonctions $\varphi : A^* \to B^*$ and $\theta : B^* \to C^*$, respectively. Let η be the function realised by $\mathcal{B} \circ \mathcal{A}$ and let $u \in A^*$. Setting $p = p_0 \cdot n$, $q = q_0 \cdot u$ and $v = q_0 * u$, one gets $\varphi(u) = n(q_0 * u)\rho(q_0 \cdot u) = nv\rho(q)$. The computation of $\eta(u)$ gives

$$\begin{aligned}\eta(u) &= m(p_0 * n)((p, q_0) * u)\omega((p, q_0) \cdot u) \\ &= m(p_0 * n)(p * (q_0 * u))\omega(p \cdot (q_0 * u), q_0 \cdot u) \\ &= m(p_0 * n)(p * v)\omega(p \cdot v, q)\end{aligned}$$

Furthermore, one has

$$\begin{aligned}\omega(p \cdot v, q) &= (p \cdot v * \rho(q))\sigma((p \cdot v) \cdot \rho(q)) \\ &= (p \cdot v * \rho(q))\sigma(p \cdot v\rho(q)) \\ &= (p \cdot v * \rho(q))\sigma(p_0 \cdot nv\rho(q)) \\ &= (p \cdot v * \rho(q))\sigma(p_0 \cdot \varphi(u))\end{aligned}$$

Coming back to the computation of $\eta(u)$, we get:

$$\begin{aligned}\eta(u) &= m(p_0 * n)(p * v)\omega(p \cdot v, q) \\ &= m(p_0 * n)(p * v)(p \cdot v * \rho(q))\sigma(p_0 \cdot \varphi(u)) \\ &= m(p_0 * n)(p_0 \cdot n * v)(p_0 \cdot nv * \rho(q))\sigma(p_0 \cdot \varphi(u)) \\ &= m(p_0 * nv\rho(q))\sigma(p_0 \cdot \varphi(u)) \\ &= m(p_0 * \varphi(u))\sigma(p_0 \cdot \varphi(u))\end{aligned}$$

$$= \theta(\varphi(u))$$

which proves that $\eta = \theta \circ \varphi$. □

Corollary 2.4. *The composition of two* [*pure*] *sequential functions is a* [*pure*] *sequential function.*

3 Sequential functions and wreath product

Since the wreath product is better defined in terms of transformation semigroups, we need to adapt two standard definitions to this setting.

First, the *transformation semigroup of a sequential transducer* is the transformation semigroup of its underlying automaton. Next, a subset L of a semigroup R is *recognised by a transformation semigroup* (P, S) if there exist a surjective morphism $\varphi\colon R \to S$, a state $p_0 \in P$ and a subset $F \subseteq P$ such that $L = \{u \in R \mid p_0 \cdot \varphi(u) \in F\}$.

Theorem 3.5. *Let $\sigma\colon A^+ \to R$ be a sequential function realised by a sequential transducer $\mathcal{T}$, and let (Q, T) be the transformation semigroup of $\mathcal{T}$. If L is a subset of R recognised by a transformation semigroup (P, S), then $\sigma^{-1}(L)$ is recognised by $(P, S) \circ (Q, T)$.*

Proof. Let $\mathcal{T} = (Q, A, R, q_0, \cdot, *, m, \rho)$. Since L is recognised by (P, S), there is a surjective morphism $\varphi\colon R \to S$, a state $p_0 \in P$ and a subset $F \subseteq P$ such that $L = \{u \in R \mid p_0 \cdot \varphi(u) \in F\}$. Let $(P, S) \circ (Q, T) = (P \times Q, W)$ and define a morphism $\psi\colon A^+ \to W$ by setting

$$(p, q) \cdot \psi(u) = (p \cdot \varphi(q * u), q \cdot u)$$

By hypothesis, the map $q \mapsto \varphi(q * u)$ is a function from Q to S and thus ψ is well defined. Let

$$I = \{(p, q) \in P \times \mathrm{Dom}(\rho) \mid p \cdot \varphi(\rho(q)) \in F\}$$

Now, since

$$(p_0 \cdot \varphi(m), q_0) \cdot \psi(u) = (p_0 \cdot \varphi(m)\varphi(q_0 * u), q_0 \cdot u)$$

one has

$$\begin{aligned}
\sigma^{-1}(L) &= \{u \in A^+ \mid \sigma(u) \in L\} = \{u \in A^+ \mid p_0 \cdot \varphi(\sigma(u)) \in F\} \\
&= \{u \in A^+ \mid p_0 \cdot \varphi(m(q_0 * u)\rho(q_0 \cdot u)) \in F\} \\
&= \{u \in A^+ \mid p_0 \cdot \varphi(m)\varphi(q_0 * u)\varphi(\rho(q_0 \cdot u)) \in F\} \\
&= \{u \in A^+ \mid (p_0 \cdot \varphi(m), q_0) \cdot \psi(u) \in I\}
\end{aligned}$$

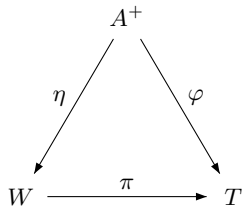
Therefore, $\sigma^{-1}(L)$ is recognised by $(P \times Q, W)$. □

4 The wreath product principle and its consequences

The aim of this section is to characterise the languages recognised by the wreath product of two transformation semigroups.

4.1 The wreath product principle

Let $X = (P, S)$ and $Y = (Q, T)$ be two transformation semigroups, let $Z = X \circ Y = (P \times Q, W)$, and let L be a language of A^+ recognised by Z. Then there exist a state (p_0, q_0), a subset F of $P \times Q$ and a morphism $\eta\colon A^+ \to W$ such that $L = \{u \in A^+ \mid (p_0, q_0)\cdot \eta(u) \in F\}$. Denote by π the natural projection from W onto T, defined by $\pi(f, t) = t$ and let $\varphi = \pi \circ \eta\colon A^+ \to T$.



Let $B = Q \times A$. Define a function $\sigma\colon A^+ \to B^+$ by

$$\sigma(a_1 a_2 \cdots a_n) = (q_0, a_1)(q_0\cdot\varphi(a_1), a_2)\cdots((q_0\cdot\varphi(a_1\cdots a_{n-1})), a_n)$$

Note that σ is a sequential function, realised by the transducer $(Q, A, B^+, q_0, \cdot, *)$ where $q\cdot a = q\cdot\varphi(a)$ and $q * a = (q, a)$.

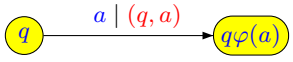


Figure 4.1. A pure sequential transducer realising σ.

We are now ready to state the wreath product principle.

Theorem 4.6 (Wreath product principle)**.** *Each language of A^+ recognised by Z is a finite union of languages of the form $U \cap \sigma^{-1}(V)$, where $U \subseteq A^+$ is recognised by Y and $V \subseteq B^+$ is recognised by X.*

Proof. First, we may assume that $F = \{(p, q)\}$ for some $(p, q) \in P \times Q$. This is a consequence of the formula

$$L = \bigcup_{(p,q)\in F} \{u \in A^+ \mid (p_0, q_0)\cdot u \in \{(p, q)\}\}$$

For each letter a, set $\eta(a) = (f_a, t_a)$. Note that $\varphi(a) = t_a$. Define a function $\alpha\colon B \to S$ by setting $\alpha(q, a) = q\cdot f_a$ and extend it to a morphism $\alpha\colon B^+ \to S$. Let $u = a_1 a_2 \cdots a_n$ be a word. Then

$$\begin{aligned}(p_0, q_0)\cdot u &= (p_0, q_0)\cdot(f_{a_1}, t_{a_1})(f_{a_2}, t_{a_2})\cdots(f_{a_n}, t_{a_n})\\ &= \big(p_0 + q_0\cdot f_{a_1} + \cdots + (q_0\cdot t_{a_1}\cdots t_{a_{n-1}})f_{a_n}, q_0\cdot t_{a_1}\cdots t_{a_n}\big)\\ &= \big(p_0 + \alpha(q_0, a_1) + \cdots + \alpha(q_0\cdot\varphi(a_1\cdots a_{n-1}), a_n), q_0\cdot\varphi(u)\big)\\ &= \big(p_0 + \alpha(\sigma(u)), q_0\cdot\varphi(u)\big)\end{aligned}$$

It follows that $(p_0, q_0)\cdot u = (p, q)$ if and only if the following two conditions are satisfied:

(1) $p_0 + \alpha(\sigma(u)) = p$,
(2) $q_0 \cdot \varphi(u) = q$.

Setting $U = \{u \in A^+ \mid q_0 \cdot \varphi(u) = q\}$ and $V = \{v \in B^+ \mid p_0 + \alpha(v) = p\}$, condition (1) can be reformulated as $u \in \sigma^{-1}(V)$, and condition (2) as $u \in U$. Thus

$$L = U \cap \sigma^{-1}(V)$$

Now, U is recognised by Y and V is recognised by X, which concludes the proof. □

We now derive a variety version of Theorem 4.6. It is stated for the case where both $\mathbf{V}$ and $\mathbf{W}$ are varieties of monoids, but similar statements hold if $\mathbf{V}$ or $\mathbf{W}$ is a variety of semigroups. Let us define the variety $\mathbf{V} * \mathbf{W}$ as the class of all divisors of wreath products of the form $S \circ T$ with $S \in \mathbf{V}$ and $T \in \mathbf{W}$.

Corollary 4.7. *Let $\mathbf{V}$ and $\mathbf{W}$ be two varieties of monoids and let $\mathcal{U}$ be the variety of languages associated with $\mathbf{V} * \mathbf{W}$. Then, for every alphabet A, $\mathcal{U}(A^*)$ is the smallest lattice containing $\mathcal{W}(A^*)$ and the languages of the form $\sigma_\varphi^{-1}(V)$, where σ_φ is the sequential function associated with a morphism $\varphi : A^* \to T$, with $T \in \mathbf{W}$ and $V \in \mathcal{V}((A \times T)^*)$.*

Proof. Since $\mathbf{W}$ is contained in $\mathbf{V} * \mathbf{W}$, $\mathcal{W}(A^*)$ is contained in $\mathcal{U}(A^*)$. Furthermore, if V and σ_φ are given as in the statement, then $\sigma_\varphi^{-1}(V) \in \mathcal{U}(A^*)$ by Theorem 3.5.

It follows from the definition of $\mathbf{V} * \mathbf{W}$ and from Proposition IV.4.26 that every language of $\mathcal{U}(A^*)$ is recognised by a wreath product of the form $S \circ T$, with $S \in \mathbf{V}$ and $T \in \mathbf{W}$. Theorem 4.6 now suffices to conclude. □

5 Applications of the wreath product principle

In this section, we give several applications of the wreath product principle. We study the operations $L \mapsto LaA^*$ and $L \mapsto La$, where a is a letter of A. Then we give a description of the languages corresponding to $\mathbf{J_1} * \mathbf{V}$, $[\![yx = x]\!] * \mathbf{V}$ and $\mathbf{r1} * \mathbf{V}$, where $\mathbf{V}$ is a variety of monoids (resp. semigroups).

5.1 The operations $T \mapsto U_1 \circ T$ and $L \mapsto LaA^*$

The study of this operation is based on the following proposition.

Proposition 5.8. *Let A be an alphabet, let $a \in A$ and let L be a language of A^* recognised by a monoid T. Then LaA^* is recognised by the wreath product $U_1 \circ T$.*

Proof. Let $\varphi\colon A^* \to T$ be a morphism recognising L, let $B = T \times A$ and let $\sigma_\varphi\colon A^* \to B^*$ be the sequential function associated with φ. Let $P = \varphi(L)$ and $C = \{(p, a) \mid p \in P\}$. Then we have

$$\begin{aligned}
\sigma_\varphi^{-1}(B^*CB^*) &= \{a_1a_2\cdots a_n \in A^* \mid \sigma_\varphi(a_1a_2\cdots a_n) \in B^*CB^*\} \\
&= \{a_1a_2\cdots a_n \in A^* \mid \text{there exists an } i \text{ such that} \\
&\qquad\qquad (\varphi(a_1a_2\cdots a_{i-1}), a_i) \in C\} \\
&= \{a_1a_2\cdots a_n \in A^* \mid \text{there exists an } i \text{ such that}
\end{aligned}$$

$$a_i = a \text{ and } a_1a_2\cdots a_{i-1} \in \varphi^{-1}(P)\}$$
$$= LaA^*$$

Since B^*CB^* is recognised by U_1, the proposition follows from Theorem 3.5. □

Proposition 5.8 leads to the following result on varieties.

Theorem 5.9. *Let* $\mathbf{V}$ *be a variety of monoids and let* $\mathcal{V}$ *be the corresponding variety. Then the variety* $\mathcal{W}$ *which corresponds to* $\mathbf{J_1} * \mathbf{V}$ *is defined as follows. For each alphabet* A, $\mathcal{W}(A^*)$ *is the Boolean algebra generated by the languages* L *and* LaA^*, *where* $a \in A$ *and* $L \in \mathcal{V}(A^*)$.

Proof. For each alphabet A, let $\mathcal{V}'(A^*)$ denote the Boolean algebra generated by the languages L and LaA^*, where $a \in A$ and $L \in \mathcal{V}(A^*)$.

We first show that $\mathcal{V}'(A^*) \subseteq \mathcal{W}(A^*)$. Since $\mathbf{J_1} * \mathbf{V}$ contains $\mathbf{V}$, $\mathcal{W}(A^*)$ contains $\mathcal{V}(A^*)$. Let $L \in \mathcal{V}(A^*)$ and $a \in A$. Then L is recognised by some monoid T of $\mathbf{V}$ and, by Proposition 5.8, LaA^* is recognised by $U_1 \circ T$. This monoid belongs to $\mathbf{J_1} * \mathbf{V}$ and hence $LaA^* \in \mathcal{W}(A^*)$.

To establish the opposite inclusion, it suffices now, by Corollary 4.7, to verify that $L \in \mathcal{V}'(A^*)$ for every language L of the form $\sigma_\varphi^{-1}(V)$, where σ_φ is the sequential function associated with a morphism of monoids $\varphi\colon A^* \to T$, with $T \in \mathbf{V}$ and V is a subset of $(T \times A)^*$ recognised by a monoid of $\mathbf{J_1}$. Let $B = T \times A$. By Proposition XIV.1.4, V is a Boolean combination of languages of the form B^*CB^*, for some subset C of $T \times A$. Since Boolean operations commute with σ_φ^{-1}, we may assume that $V = B^*CB^*$, where $C = (t, a)$ for some $(t, a) \in B$. In this case

$$\begin{aligned} L &= \{u \in A^* \mid \sigma_\varphi(u) \in B^*CB^*\} \\ &= \{a_1a_2\cdots a_n \in A^* \mid (\varphi(a_1\cdots a_{i-1}), a_i) = (t, a) \text{ for some } i\} \\ &= \varphi^{-1}(t)aA^* \end{aligned} \tag{5.1}$$

Now, $\varphi^{-1}(t)$ is recognised by T and thus $\varphi^{-1}(t) \in \mathcal{V}(A^*)$. It follows that $\varphi^{-1}(t)aA^* \in \mathcal{V}'(A^*)$ and thus $L \in \mathcal{V}'(A^*)$. Therefore $\mathcal{W}(A^*) \subseteq \mathcal{V}'(A^*)$. □

5.2 The operations $T \mapsto \bar{\mathbf{2}} \circ T$ and $L \mapsto La$

Recall that $\bar{\mathbf{2}}$ denotes the transformation semigroup $(\{1, 2\}, \{1, 2\})$ with the action defined by $r \cdot s = s$. The results are quite similar to those presented in Section 5.1, but the monoid U_1 is now replaced by $\bar{\mathbf{2}}$.

Proposition 5.10. *Let* A *be an alphabet, let* $a \in A$ *and let* L *be a language of* A^* *recognised by a monoid* T. *Then* La *is recognised by the wreath product* $\bar{\mathbf{2}} \circ T$.

Proof. Let $\varphi : A^* \to T$ be a morphism recognising L, let $B = T \times A$ and let $\sigma_\varphi\colon A^* \to B^*$ be the sequential function associated with φ. Let $P = \varphi(L)$ and $C = \{(p, a) \mid p \in P\}$. Then we have

$$\begin{aligned} \sigma_\varphi^{-1}(B^*C) &= \{u \in A^* \mid \sigma_\varphi(u) \in B^*C\} \\ &= \{a_1a_2\cdots a_n \in A^+ \mid (\varphi(a_1a_2\cdots a_{n-1}), a_n) \in C\} \\ &= \{a_1a_2\cdots a_n \in A^+ \mid a_n = a \text{ and } a_1a_2\cdots a_{n-1} \in \varphi^{-1}(P)\} \end{aligned}$$

$$= La$$

Since the language B^*C is recognised by $\bar{\mathbf{2}}$, the proposition now follows from Theorem 3.5. □

A result similar to Proposition 5.10 holds if T is a semigroup which is not a monoid. The proof is left as an exercise to the reader.

Proposition 5.11. *Let A be an alphabet, let $a \in A$ and let L be a language of A^+ recognised by an semigroup T such that $T \neq T^1$. Then the languages La and $\{a\}$ are recognised by the wreath product $\bar{\mathbf{2}} \circ (T^1, T)$.*

We now turn to varieties. Observe that the semigroup $\bar{\mathbf{2}}$ generates the variety $[\![yx = x]\!]$.

Theorem 5.12. *Let $\mathbf{V}$ be a variety of monoids and let $\mathcal{V}$ be the corresponding variety. Let $\mathcal{W}$ be the variety corresponding to $[\![yx = x]\!] * \mathbf{V}$. Then, for each alphabet A, $\mathcal{W}(A^+)$ is the lattice generated by the languages L (contained in A^+) or La, where $a \in A$ and $L \in \mathcal{V}(A^*)$.*

Proof. For each alphabet A, let $\mathcal{V}'(A^+)$ be the lattice generated by the languages L (contained in A^+) or La, where $a \in A$ and $L \in \mathcal{V}(A^*)$.

We first show that $\mathcal{V}'(A^+) \subseteq \mathcal{W}(A^+)$. Since $[\![yx = x]\!] * \mathbf{V}$ contains $\mathbf{V}$, every language of $\mathcal{V}(A^*)$ contained in A^+ is also in $\mathcal{W}(A^+)$. Let $L \in \mathcal{V}(A^*)$ and $a \in A$. Then L is recognised by some monoid T of $\mathbf{V}$ and, by Proposition 5.10, La is recognised by $\bar{\mathbf{2}} \circ T$. Now this semigroup belongs to $[\![yx = x]\!] * \mathbf{V}$ and thus $La \in \mathcal{W}(A^+)$.

To establish the opposite inclusion, it suffices now, by Corollary 4.7, to verify that $L \in \mathcal{V}'(A^+)$ for every language L of the form $\sigma_\varphi^{-1}(V)$, where σ_φ is the sequential function associated with a morphism of monoids $\varphi\colon A^* \to T$, with $T \in \mathbf{V}$, and V is a language of $(T \times A)^+$ recognised by a transformation semigroup of $[\![yx = x]\!]$. Let $B = T \times A$. Then V is a finite union of languages of the form B^*c, for some letter $c \in B$. Since union commutes with σ_φ^{-1}, we may assume that $V = B^*c$, where $c = (t, a)$ for some $(t, a) \in B$. In this case

$$\begin{aligned} L &= \{u \in A^+ \mid \sigma_\varphi(u) \in B^*c\} \\ &= \{a_1 a_2 \cdots a_n \in A^+ \mid (\varphi(a_1 \cdots a_{n-1}), a_n) = (t, a)\} \\ &= \varphi^{-1}(t)a \end{aligned}$$

Now, $\varphi^{-1}(t)$ is recognised by T and thus $\varphi^{-1}(t) \in \mathcal{V}(A^*)$. It follows that $\varphi^{-1}(t)a \in \mathcal{V}'(A^+)$ for each letter a and thus $L \in \mathcal{V}'(A^+)$. Therefore $\mathcal{W}(A^+) \subseteq \mathcal{V}'(A^+)$. □

The semigroup version of Theorem 5.12 can be obtained by using Proposition 5.11 instead of Proposition 5.10. Recall that $B(1, 2)$ denotes the semigroup $\{a, b\}$ equipped the multiplication defined by $aa = ba = a$ and $ab = bb = b$.

Theorem 5.13. *Let $\mathbf{V}$ be a variety of semigroups which is not a variety of groups, and let $\mathcal{V}$ be the corresponding variety. Let $\mathcal{W}$ be the variety corresponding to $[\![yx = x]\!] * \mathbf{V}$. Then, for each alphabet A, $\mathcal{W}(A^+)$ is the Boolean algebra generated by the languages L, $\{a\}$ or La, where $a \in A$ and $L \in \mathcal{V}(A^+)$.*

Recall that the variety $\mathrm{r}\mathbf{1}$ is defined by the identity $yx^\omega = x^\omega$. It is known to be the smallest variety of semigroups containing $B(1,2)$ and closed under semidirect product [42]. Therefore, Theorems 5.12 and 5.13 lead to a language interpretation of the operation $\mathbf{V} \to \mathrm{r}\mathbf{1} * \mathbf{V}$.

Corollary 5.14. *Let $\mathbf{V}$ be a variety of monoids and let $\mathcal{V}$ be the corresponding positive variety. Let $\mathcal{W}$ be the positive variety corresponding to $\mathrm{r}\mathbf{1} * \mathbf{V}$. Then, for each alphabet A, $\mathcal{W}(A^+)$ is the smallest lattice containing $\mathcal{V}(A^*)$ and closed under the operation $L \mapsto Lu$, for each $u \in A^*$.*

Corollary 5.15. *Let $\mathbf{V}$ be a variety of ordered semigroups which is not a variety of groups and let $\mathcal{V}$ be the corresponding positive variety. Let $\mathcal{W}$ be the positive variety corresponding to $\mathrm{r}\mathbf{1} * \mathbf{V}$. Then, for each alphabet A, $\mathcal{W}(A^+)$ is the smallest lattice containing $\mathcal{V}(A^+)$ and the languages $\{a\}$, for each $a \in A$, and closed under the operation $L \mapsto Lu$, for each $u \in A^*$.*

5.3 The operation $T \mapsto U_2 \circ T$ and star-free expressions

Recall that U_2 denotes the monoid $\{1, a_1, a_2\}$ defined by $a_1a_1 = a_2a_1 = a_1$ and $a_1a_2 = a_2a_2 = a_2$.

Proposition 5.16. *Let A be an alphabet and let T be a monoid. Then every language of A^* recognised by $U_2 \circ T$ is a Boolean combination of languages of the form K or $Ka(LbA^*)^c$ where $a, b \in A$ and K and L are recognised by T.*

Proof. Let L be a language of A^* recognised by $U_2 \circ T$ and let $B = T \times A$. The wreath product principle tells us that L is a finite union of languages of the form $U \cap \sigma^{-1}(V)$, where $U \subseteq A^*$ is recognised by T and $V \subseteq B^*$ is recognised by U_2. By Proposition XIV.1.7, every language of B^* recognised by U_2 is a Boolean combination of languages of the form B^*bC^*, where $b \in B$ and $C \subseteq B$. Therefore, it suffices to prove that a language of the form $\sigma^{-1}(V)$ has the right form. We claim that

$$\sigma^{-1}(B^*bC^*) = \varphi^{-1}(m)a\Bigg(\bigcup_{(n,c)\notin C} \varphi^{-1}\Big(\big(m\varphi(a)\big)^{-1}n\Big)cA^*\Bigg)^c \tag{5.2}$$

Indeed, let $b = (t, a)$ and let $u = a_1a_2\cdots a_n$ be a word of A^*. Then $\sigma(u)$ belongs to B^*bC^* if and only if there exists an i such that $\varphi(a_1a_2\cdots a_{i-1}) = m$, $a_i = a$ and, for $i \leqslant j \leqslant n-1$, $(\varphi(a_1\cdots a_j), a_{j+1}) \in C$. The negation of the latter condition can be stated as follows: there is a j such that $(\varphi(a_1\cdots a_j), a_{j+1}) \notin C$. In other words, $\varphi(a_1\cdots a_j) = n$ and $a_{j+1} = c$ for some $(n, c) \notin C$. Now, if $\varphi(a_1a_2\cdots a_{i-1}) = m$, $a_i = a$ and $\varphi(a_1\cdots a_j) = n$, then $m\varphi(a)\varphi(a_{i+1}\cdots a_j) = n$ and therefore $\varphi(a_{i+1}\cdots a_j) \in \big(m\varphi(a)\big)^{-1}n$ and $a_{i+1}\cdots a_n \in \varphi^{-1}\Big(\big(m\varphi(a)\big)^{-1}n\Big)cA^*$. This justifies (5.2).

Now, each language $\varphi^{-1}(m)$ and $\varphi^{-1}\Big(\big(m\varphi(a)\big)^{-1}n\Big)$ is recognised by T, which concludes the proof. □

Proposition 5.16 leads to a new proof of the fact that a language recognised by an aperiodic monoid is star-free, the difficult part of Theorem VI.3.3. Proposition 5.16 shows that if all the languages recognised by T are star-free, then

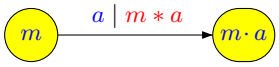
$U_2 \circ T$ has the same property. Now Theorem XVIII.3.15 shows that every aperiodic monoid divides a wreath product of copies of U_2 and it suffices to proceed by induction to conclude.

6 Exercises

Exercise 1. Let L be a recognisable language on the alphabet A, a a letter of A and u a word of A^*. Let $N(u)$ denote the number of factorisations of u of the form $u = u_0 a u_1$ with $u_0 \in L$ et $u_1 \in A^*$.

(1) Let $A = \{a, b\}$ and $L = (ab)^*$. Compute $N(u)$ when u is one of the following words: ab, $abab$, $ababab$, $ababbab$.

(2) Find a sequential transducer computing $N(u)$. **Hint**: one can modify the algorithm used to convert the syntactic monoid of L to a finite deterministic automaton (see Proposition IV.3.20) and supplement this automaton with an output function taking values in the monoid $(\mathbb{N}, +)$. In other words, which values should one take for each $m * a$ in order to have $1 * u = N(u)$ for all $u \in A^*$?



Exercise 2. Let r and n be two integers such that $0 \leqslant r < n$ and let

$$(LaA^*)_{r,n} = \{u \in A^* \mid N(u) \equiv r \bmod n\}$$

Deduce from Exercise 1 that $(LaA^*)_{r,n}$ is recognised by the wreath product $\mathbb{Z}/n\mathbb{Z} \circ M$.

Chapter XX

Concatenation hierarchies

In the seventies, several classification schemes for the rational languages were proposed, based on the alternate use of certain operators (union, complementation, product and star). Some forty years later, although much progress has been done, several of the original problems are still open. Furthermore, their significance has grown considerably over the years, on account of the successive discoveries of links with other fields, like non commutative algebra [42], finite model theory [168], structural complexity [17] and topology [83, 101, 104].

Roughly speaking, the concatenation hierarchy of a given class of recognisable languages is built by alternating Boolean operations (union, intersection, complement) and polynomial operations (union and marked product). For instance, the *Straubing-Thérien hierarchy* [166, 156, 158] is based on the empty and full languages of A^* and the *group hierarchy* is built on the group-languages, the languages recognised by a finite permutation automaton. It can be shown that, if the basis of a concatenation hierarchy is a variety of languages, then every level is a positive variety of languages [8, 9, 118], and therefore corresponds to a variety of finite ordered monoids [101]. These varieties are denoted by $\mathbf{V}_n$ for the Straubing-Thérien hierarchy and $\mathbf{G}_n$ for the group hierarchy.

In Section 2, we describe the hierarchy of the first-order logic of the linear order, corresponding to the alternate use of existential and universal quantifiers. We show that this hierarchy corresponds, on words, to the concatenation hierarchy described in Section 1. This leads to a doubly satisfying situation: first-order formulas not only correspond globally to a natural class of recognisable sets (the star-free sets), but this correspondence holds level by level.

1 Concatenation hierarchies

In this section, we introduce a hierarchy among star-free languages of A^*, known as the *Straubing-Thérien's hierarchy*, or *concatenation hierarchy*.[1] For historical reasons, this hierarchy is indexed by half-integers. The level 0 consists of the languages $\emptyset$ and A^*. The other levels are defined inductively as follows:

(1) the level $n + 1/2$ is the class of union of marked products of languages of level n;

[1] A similar hierarchy, called the *dot-depth hierarchy* was previously introduced by Brzozowski, but the Straubing-Thérien's hierarchy is easier to define.

(2) the level $n+1$ is the class of Boolean combination of languages of marked products of level n.

We call the levels n (for some nonegative integer n) the *full levels* and the levels $n + 1/2$ the *half levels*.

In particular level $1/2$ consists of the shuffle ideals studied in Section VII.2. Recalll that a shuffle ideal is a finite union of languages of the form

$$A^* a_1 A^* a_2 A^* \cdots A^* a_k A^*$$

where $a_1, \ldots, a_k \in A$. Level 1 consists of the piecewise testable languages studied in Section VII.3.

It is not clear at first sight whether the Straubing-Thérien's hierarchy does not collapse, but this question was solved in 1978 by Brzozowski and Knast [22].

Theorem 1.1. *The Straubing-Thérien's hierarchy is infinite.*

It is a major open problem on regular languages to know whether one can decide whether a given star-free language belongs to a given level.

Problem. *Given a half integer n and a star-free language L, decide whether L belongs to level n.*

One of the reasons why this problem is particularly appealing is its close connection with finite model theory, presented in Chapter IX.

Only few decidability results are known. It is obvious that a language has level 0 if and only if its syntactic monoid is trivial. It was shown in Theorem VII.2.10 that a language has level $1/2$ if and only if its syntactic ordered monoid satisfies the identity $1 \leqslant x$. Theorem VII.3.13, due to I. Simon [145] states that a language has level 1 if and only if its syntactic monoid is finite and $\mathcal{J}$-trivial. The decidability of level $3/2$ was first proved by Arfi [8, 9] and the algebraic characterisation was found by Pin-Weil [115]. It relies on the following result

Theorem 1.2. *A language is of level $3/2$ if and only if its syntactic ordered monoid belongs to the Mal'cev product $[\![x^\omega y x^\omega \leqslant x^\omega]\!] \mathbin{Ⓜ} [\![x^2 = x, xy = yx]\!]$. This condition is decidable.*

The decidability of levels 2 and $5/2$ was recently proved by Place and Zeitoun [127] and the decidability of level $7/2$ was proved by Place [123].

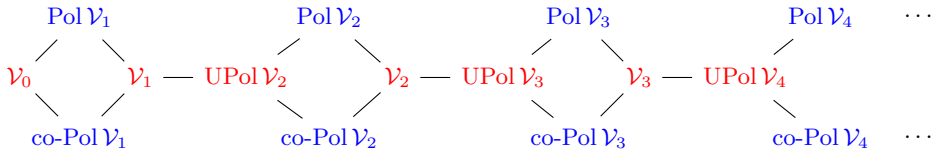


Figure 1.1. The Straubing-Thérien hierarchy.

2 Logical hierarchy

We shall see in this section how to refine the characterisation of the first-order definable sets of the logic of the linear order (Theorem IX.4.12). We shall actually establish a bijection between the concatenation hierarchy of star-free sets and the logical hierarchy $\Sigma_n[<]$. Since the concatenation hierarchy is infinite, it will show that the hierarchy defined on formulas by quantifier alternation is also infinite.

One can set up a hierarchy inside first-order formulas as follows. Let $\Sigma_0 = \Pi_0$ be the set of quantifier-free formulas. Next, for each $n \geqslant 0$, denote by Σ_{n+1} the least set Δ of formulas such that

(1) Δ contains all Boolean combinations of formulas of Σ_n,
(2) Δ is closed under finite disjunctions and conjunctions,
(3) if $\varphi \in \Delta$ and if x is a variable, $\exists x\varphi \in \Delta$.

Similarly, let Π_{n+1} denote the least set Γ of formulas such that

(1) Γ contains all Boolean combinations of formulas of Π_n,
(2) Γ is closed under finite disjunctions and conjunctions,
(3) if $\varphi \in \Gamma$ and if x is a variable, $\forall x\varphi \in \Gamma$.

In particular, Σ_1 is the set of *existential* formulas, that is, formulas of the form

$$\exists x_1\ \exists x_2\ \dots\ \exists x_k\ \varphi$$

where φ is quantifier-free.

Finally, we let $\mathcal{B}\Sigma_n$ denote the set of Boolean combinations of formulas of Σ_n.

Proposition IX.2.2 can be improved to take into account the level of the formula in the Σ_n hierarchy.

Proposition 2.3. *For each integer $n \geqslant 0$,*

(1) *Every formula of Σ_n is logically equivalent to a formula in prenex normal form in which the quantifier prefix is a sequence of n (possibly empty) alternating blocks of existential and universal quantifiers, starting with a block of existential quantifiers.*
(2) *Every formula of Π_n is logically equivalent to a formula in prenex normal form in which the quantifier prefix is a sequence of n (possibly empty) alternating blocks of existential and universal quantifiers, starting with a block of universal quantifiers.*

For example, the formula

$$\underbrace{\exists x_1\ \exists x_2\ \exists x_3}_{\text{block 1}}\ \underbrace{\forall x_4\ \forall x_5}_{\text{block 2}}\ \underbrace{\exists x_6\ \exists x_7}_{\text{block 3}}\ \varphi(x_1, \dots, x_6)$$

belongs to Σ_3 (and also to all Σ_n with $n \geqslant 3$). Similarly the formula

$$\underbrace{}_{\text{block 1}}\ \underbrace{\forall x_4\ \forall x_5}_{\text{block 2}}\ \underbrace{\exists x_6\ \exists x_7}_{\text{block 3}}\ \varphi(x_1, \dots, x_7)$$

belongs to Σ_3 and to Π_2, but not to Σ_2, since the counting of blocks of a Σ_n-formula should always begin by a possibly empty block of existential quantifiers.

Theorem 2.4. *For each integer $n \geqslant 0$,*

(1) *A language is definable by a formula of $\mathcal{B}\Sigma_n[<]$ if and only if it is a star-free set of level n.*

(2) *A language is definable by a formula of $\Sigma_{n+1}[<]$ if and only if it is a star-free set of level $n + 1/2$.*

Proof. The first part of the proof consists in converting sets of words to formulas. To start with, the equalities

$$L(\mathbf{true}) = A^* \qquad L(\mathbf{false}) = \emptyset$$

show that each star-free subset of level 0 is definable by a formula of $\mathcal{B}\Sigma_0$.

By induction, suppose that the star-free sets of level $n + 1/2$ are definable by a formula of $\Sigma_{n+1}[<]$. Proposition IX.3.8 shows now that the star-free sets of level $n + 1$ are definable by a formula of $\mathcal{B}\Sigma_{n+1}[<]$.

Suppose now that the star-free sets of level n are definable by a formula of $\mathcal{B}\Sigma_n[<]$. If X is a star-free subset of A^* of level $n + 1/2$, X can be written as a finite union of subsets of the form

$$X_0 a_1 X_1 a_2 \cdots a_k X_k$$

where $k \geqslant 0$, X_0, X_1, ..., X_k are star-free sets of level n of A^* and a_1, ..., a_k are letters.

We now use again the "relativised" version of the first-order formula given by Proposition IX.2.4. If $X_0 = L(\varphi_0)$, $X_1 = L(\varphi_1)$, ..., $X_k = L(\varphi_k)$, we have $X_0 a_1 X_1 \cdots a_k X_k = L(\varphi)$, where φ is the formula

$$\exists x_1 \cdots \exists x_k \bigwedge_{1 \leqslant i \leqslant k-1} (x_i < x_{i+1}) \bigwedge_{1 \leqslant i \leqslant k} \mathbf{a}_i(x_i) \wedge \varphi_1(\min, x_1 - 1) \bigwedge_{1 \leqslant i \leqslant k-1} \varphi(x_i + 1, x_{i+1} - 1) \wedge \varphi(x_k, \max)$$

showing that $X_0 a_1 X_1 \cdots a_k X_k$ is definable by a formula of Σ_{n+1}. It follows, by Proposition IX.3.8, that if the star-free subsets of level n are definable by a formula of $\mathcal{B}\Sigma_n[<]$, the star-free subsets of level $n + 1/2$ are definable by a formula of $\Sigma_{n+1}[<]$.

We now arrive at the second part of the proof, the conversion of formulas to star-free sets. We recall the scheme of the proof of Theorem IX.4.12, which consists in arguing by induction on the formation rules of formulas. For this purpose, we build, for each alphabet B, a hierarchy of star-free subsets of B^* which differs from the usual hierarchy, in particular at level 0.

The induction that follows forces very precise requirements on the subsets of B^* of level 0. In particular, they have to contain the subsets of the form $L(\mathbf{a}(x_i))$, $L(x_i = x_j)$ and $L(x_i < x_j)$ and to be closed under quotient. This leads us to the following definition, which will be subsequently justified by the lemmas 2.6 and 2.7. For each $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_p) \in \{0,1\}^p$, we set

$$K(\varepsilon) = \{u \in B_p^* \mid \text{for } 1 \leqslant i \leqslant p,\ u_i \text{ contains exactly } \varepsilon_i \text{ occurrences of } 1\}$$

Note that $K_p = K(1, \ldots, 1)$. For $a \in A$, $i \in \{1, \ldots, p\}$ and $\varepsilon \in \{0,1\}^p$, we set

$$K(a, i, \varepsilon) = \{u \in K(\varepsilon) \mid \text{if } u_{i,n_i} = 1, \text{ then } u_{0,n_i} = a\}$$

This definition is interesting only if $\varepsilon_i = 1$. In this case, $K(a, i, \varepsilon)$ is the set of the words of $K(\varepsilon)$ such that, if n_i is the unique position of 1 in u_i, the letter of u_0 in position n_i is an a. Finally, if $i, j \in \{1, \ldots, p\}$, we set

$$K_=(i, j, \varepsilon) = \{u \in K(\varepsilon) \mid u_{i,n} = 1 \text{ if and only if } u_{j,n} = 1\}$$
$$K_<(i, j, \varepsilon) = \{u \in K(\varepsilon) \mid \text{if } u_{i,n_i} = 1 \text{ and } u_{j,n_j} = 1, \text{ then } n_i < n_j\}$$

Again, these definitions are only interesting if $\varepsilon_i = \varepsilon_j = 1$. In this case, $K_=(i, j, \varepsilon)$ $[K_<(i, j, \varepsilon)]$ is the set of the words of $K(\varepsilon)$ such that the unique position of 1 in u_i is equal to [precedes] the unique position of 1 in u_j.

The languages of B_p^* of level 0 form the least set of subsets closed under finite intersection and finite union and containing the subsets of the form $K(\varepsilon)$, $K(a, i, \varepsilon)$, $K_=(i, j, \varepsilon)$ and $K_<(i, j, \varepsilon)$ for each $\varepsilon \in \{0, 1\}^p$, $a \in A$ and $i, j \in \{1, \ldots, p\}$.

Next, level $n + 1/2$ is the least set of subsets closed under finite intersection and finite union and containing the subsets of the form

$$L_0 b_1 L_1 b_2 \cdots b_k L_k$$

where $k \geqslant 0$, $b_1, \ldots, b_k \in B_p$ and L_0, L_1, $\ldots$, L_k are subsets of level n of B_p^*. The level $n+1$ is the closure of level $n+1/2$ under finite union, finite intersection and the operations $X \to Y - X$ where Y is a subset of level 0.

For $p = 0$, we recover the usual hierarchy.

Lemma 2.5. *The hierarchy of subsets of B_0^* coincides with the concatenation hierarchy.*

Proof. We have $B_0 = A$ and the only subsets of level 0 are $\emptyset$ and A^*. It follows that the two hierarchies coincide. □

The subsets of level n $[n + 1/2]$ are closed under quotients.

Lemma 2.6. *Let $n \geqslant 0$ and let X be a language of B_p^* of level n $[n+1/2]$. For each $u, v \in B_p^*$, the languages $u^{-1}X$ and Xv^{-1} are of level n $[n + 1/2]$.*

Proof. Arguing by induction on the length of u, it suffices to treat the case where $u = b$, with $b \in B_p$. We treat only the case of a right quotient, the other case being dual.

We already know that quotients commute with Boolean operations: if X_1 and X_2 are languages of B_p^*, then

$$(X_1 \cup X_2)b^{-1} = X_1 b^{-1} \cup X_2 b^{-1}$$
$$(X_1 \cap X_2)b^{-1} = X_1 b^{-1} \cap X_2 b^{-1}$$
$$(X_1 - X_2)b^{-1} = X_1 b^{-1} - X_2 b^{-1}$$

Moreover, if $b_1, \ldots, b_k$ are letters of B_p, L_0, L_1, $\ldots$, L_k are languages of B_p^*, one has the formula

$$(L_0 b_1 L_1 b_2 \cdots b_k L_k)b^{-1} =$$

$$\begin{cases} L_0b_1L_1b_2\cdots b_k(L_kb^{-1}) \cup L_0b_1L_1b_2\cdots b_{k-1}L_{k-1} & \text{if } 1 \in L_k \text{ and } b = b_k \\ L_0b_1L_1b_2\cdots b_k(L_kb^{-1}) & \text{otherwise} \end{cases} \tag{2.1}$$

We now conclude the proof by induction. The subsets of level 0 are closed under quotient. Indeed let K' be one of the subsets $K'(\varepsilon)$, $K'(a,i,\varepsilon)$, $K'_=(i,j,\varepsilon)$ or $K'_<(i,j,\varepsilon)$. If u is not a suffix of any word of K', then of course $K'u^{-1} = \emptyset$. Otherwise, we have in particular, denoting by ε'_i the number of occurrences of 1 in u_i, $\varepsilon'_i \leqslant \varepsilon_i$ and

$$\begin{aligned} K'(\varepsilon)u^{-1} &= K'(\varepsilon - \varepsilon') \\ K'(a,i,\varepsilon)u^{-1} &= K'(a,i,\varepsilon - \varepsilon') \\ K'_=(i,j,\varepsilon)u^{-1} &= K'_=(i,j,\varepsilon - \varepsilon') \\ K'_<(i,j,\varepsilon)u^{-1} &= K'_<(i,j,\varepsilon - \varepsilon') \end{aligned}$$

If the subsets of level n are closed under quotient, Formula 2.1 shows that every quotient of a marked product of such subsets is of level $n + 1/2$. It follows that the subsets of level $n + 1/2$ are closed under quotients since quotients commute with Boolean operations. Next, the same argument shows that the subsets of level $n + 1$ are closed under quotients, completing the induction. □

We come back to the conversion of formulas to star-free sets. The induction starts with the formulas of $\mathcal{B}\Sigma_0$.

Lemma 2.7. *If $\varphi \in \mathcal{B}\Sigma_0$, $S_p(\varphi)$ is a subset of level 0 of B_p^*.*

Proof. We may assume that φ is in disjunctive normal form. Then we get rid of the negations of atomic formulas by replacing

$$\begin{aligned} \neg(x = y) \quad &\text{by} \quad (x < y) \vee (y < x) \\ \neg(x < y) \quad &\text{by} \quad (x = y) \vee (y < x) \\ \neg\mathbf{a}(x) \quad &\text{by} \quad \bigvee_{b \neq a} \mathbf{b}(x) \end{aligned}$$

Thus φ is now a disjunction of conjunctions of atomic formulas, and by Proposition IX.3.8, we are left with the case of atomic formulas. Finally, the formulas

$$\begin{aligned} S_p(\mathbf{a}(x_i)) &= K(a,i,1,\dots,1) \\ S_p(x_i = x_j) &= K_=(i,j,1,\dots,1) \\ S_p(x_i < x_j) &= K_<(i,j,1,\dots,1) \end{aligned}$$

conclude the proof since these subsets are of level 0 by definition. □

The key step of the induction consists to convert existential quantifiers to concatenation products.

Lemma 2.8. *Let φ be a formula of $\mathcal{B}\Sigma_n$ and $x_{i_1}, \dots, x_{i_k}$ be free variables of φ. If $S_p(\varphi)$ is a subset of level n of B_p^*, then $S_{p-k}(\exists x_{i_1} \cdots \exists x_{i_k} \varphi)$ is a subset of level $n + 1/2$ of B_p^*.*

Proof. Up to a permutation of the free variables, we may assume that $x_{i_1} = x_1$, $\ldots$, $x_{i_k} = x_k$. We first establish the formula

$$S_{p-k}(\exists x_1 \cdots \exists x_k \varphi) = \pi(S_p(\varphi)) \tag{2.2}$$

where $\pi : B_p \to B_{p-k}$ is the function erasing the components with index 1 to k, defined by:

$$\pi_i(b_0, b_1, \ldots, b_p) = (b_0, b_{k+1}, \ldots, b_p)$$

Indeed, if $u = (u_0, u_1, \ldots, u_p) \in S_p(\varphi)$, we have in particular $u \in K_p$. For $1 \leqslant i \leqslant p$, let n_i denote the position of the unique 1 of the word u_i. Then we have $u_0 \models \varphi[\nu]$, where ν is the valuation defined by $\nu(x_i) = n_i$ for $1 \leqslant i \leqslant p$. Now $\nu = \nu'[\binom{n_1}{x_1} \cdots \binom{n_k}{x_k}]$, where ν' is the valuation defined by $\nu'(x_i) = n_i$ for $k+1 \leqslant i \leqslant p$. It follows $u_0 \models (\exists x_1 \cdots \exists x_k \varphi)[\nu']$ and as $\pi(u)_0 = u_0$, it follows $\pi(u) \in S_{p-k}(\exists x_1 \cdots \exists x_k \varphi)$.

Conversely, if $v \in S_{p-k}(\exists x_1 \cdots \exists x_k \varphi)$, there exist positions n_1, $\ldots$, n_k such that

$$v_0 \models \varphi[\nu'[\tbinom{n_1}{x_1} \cdots \tbinom{n_k}{x_k}]].$$

Let u be the unique word of K_p such that $\pi(u) = v$ and such that, for $1 \leqslant i \leqslant k$, the position of the unique 1 of u_i is n_i. Then $u_0 = v_0$, and as $\nu = \nu'[\binom{n_1}{x_1} \cdots \binom{n_k}{x_k}]$, we have $u_0 \models \varphi[\nu]$, whence $u \in S_p(\varphi)$, proving Formula 2.2.

We shall now express $S_p(\varphi)$ as a finite union of marked products of subsets of level n. Set $L = S_p(\varphi)$, $D = \{\ b \in B_p \mid \text{for } 1 \leqslant i \leqslant k,\ b_i = 0\ \}$ and, for $I \subseteq \{1, \ldots, p\}$,

$$C_I = \{\ b \in B_p \mid \text{for each } i \in I\ b_i = 1\ \}$$

If $u \in L$, there exists, for $1 \leqslant i \leqslant k$, a unique position n_i such that $u_{i,n_i} = 1$. Then there exists an ordered partition $P = (I_1, \ldots, I_r)$ of $\{1, \ldots, k\}$ such that u is uniquely factored as $u = u_0 b_1 u_1 b_2 \cdots b_r u_r$, where $b_j \in C_{I_j}$, $u_0, \ldots, u_r \in D^*$.

		a_1			a_2				a_3		$\cdots$
0	0	1	0	0	0	0	0	0	0	0	$\cdots$
0	0	0	0	0	0	0	0	0	1	0	$\cdots$
0	0	0	0	0	1	0	0	0	0	0	$\cdots$
0	0	1	0	0	0	0	0	0	0	0	$\cdots$
					$\vdots$						$\cdots$

Next we obtain the formula below, in which $Q(x) = \{u \in B_p^* \mid u^{-1}L = x^{-1}L\}$ and $\mathcal{P}$ denotes the set of ordered partitions of $\{1, \ldots, k\}$:

$$L = \bigcup_{P \in \mathcal{P}} \bigcup_{(x_0, b_1, x_1, \ldots, b_r, x_r) \in E_P} [Q(x_0)]\, b_1\, [(x_0 b_1)^{-1} Q(x_0 b_1 x_1)]\, b_2 \cdots \cdots b_r\, [(x_0 b_1 x_1 \cdots b_r)^{-1} Q(x_0 b_1 x_1 \cdots b_r x_r)] \tag{2.3}$$

in which if $P = (I_1, \ldots, I_r)$ is an ordered partition, E_P is the set of

$$(x_0, b_1, x_1, \ldots, b_r, x_r)$$

such that $x_0 b_1 x_1 \cdots b_r x_r \in L$, with $x_0, x_1, \ldots, x_p \in B_p^*$ and, for $1 \leqslant j \leqslant r$, $b_j \in C_{I_j}$.

Let L' be the right-hand side of the formula 2.3. First of all, if $u \in L$, there exists an ordered partition $(I_1, \ldots, I_r)$ of $\{1, \ldots, k\}$ and a factorisation $u = u_0 b_1 u_1 \cdots b_r u_r$, with, for $1 \leqslant j \leqslant r$, $b_j \in C_{I_j}$. Now, by construction, $x \in Q(x)$ for each x, and hence

$$u_0 \in Q(u_0),\ u_1 \in (u_0 b_1)^{-1} Q(u_0 b_1 u_1),\ \ldots,\\ u_r \in (u_0 b_1 \cdots u_{r-1} b_r)^{-1} Q(u_0 b_1 \cdots u_{r-1} b_r u_r)$$

thereby establishing the inclusion $L \subseteq L'$. Conversely, if there exists an ordered partition P of $\{1, \ldots, k\}$ and if

$$u \in [Q(x_0)]\, b_1\, [(x_0 b_1)^{-1} Q(x_0 b_1 x_1)]\, b_2 \cdots\\ b_k\, [(x_0 b_1 x_1 \cdots b_r)^{-1} Q(x_0 b_1 x_1 \cdots b_r x_r)]$$

for some $(x_0, b_1, x_1, \ldots, b_k, x_r) \in E_P$, we have $u = u_0 b_1 u_1 \cdots b_r u_r$, with $b_j \in C_{I_j}$ and

$$u_0 \in Q(x_0), u_1 \in (x_0 b_1)^{-1} Q(x_0 b_1 x_1), \ldots,\\ u_r \in (x_0 b_1 x_1 \cdots b_r)^{-1} Q(x_0 b_1 x_1 \cdots b_r x_r)$$

Therefore, for $0 \leqslant i \leqslant r$, $x_0 b_1 x_1 \cdots b_i u_i \in Q(x_0 b_1 x_1 \cdots b_i x_i)$, whence

$$(x_0 b_1 x_1 \cdots b_i u_i)^{-1} L = (x_0 b_1 x_1 \cdots b_i x_i)^{-1} L$$

Next we show that $x_0 b_1 x_1 \cdots b_i x_i b_{i+1} u_{i+1} \cdots b_r u_r \in L$ by induction on $r - i$. For $i = r$, the result follows from the fact that $(x_0, b_1, x_1, \ldots, b_r, x_r) \in E_P$. Next, if

$$x_0 b_1 x_1 \cdots b_i x_i b_{i+1} u_{i+1} \cdots b_r u_r \in L$$

we have $b_{i+1} u_{i+1} \cdots b_r u_r \in (x_0 b_1 x_1 \cdots b_i x_i)^{-1} L$ and hence

$$b_{i+1} u_{i+1} \cdots b_r u_r \in (x_0 b_1 x_1 \cdots b_i u_i)^{-1} L$$

whence $x_0 b_1 x_1 \cdots b_{i-1} x_{i-1} b_i u_i \cdots b_r u_r \in L$, allowing the induction. Consequently

$$u = u_0 b_1 u_1 \cdots b_r u_r \in L$$

and the inclusion $L' \subseteq L$ is proved.

A second formula gives an expression of $Q(x)$ in terms of quotients of L.

$$Q(x) = \Big(\bigcap_{y \in x^{-1} L} L y^{-1} \Big) - \Big(\bigcup_{y \notin x^{-1} L} L y^{-1} \Big) \tag{2.4}$$

Let indeed $Q'(x)$ be the right-hand side of (2.4). If $u \in Q(x)$, then $u^{-1} L = x^{-1} L$. Therefore the following sequence of equivalences hold

$$y \in x^{-1} L \iff y \in u^{-1} L \iff uy \in L \iff u \in L y^{-1}$$

showing that $Q(x) \subseteq Q'(x)$. Let now $u \in Q'(x)$. If $y \in x^{-1} L$, then $u \in L y^{-1}$ and hence $uy \in L$ and $y \in u^{-1} L$. Similarly, if $y \notin x^{-1} L$, then $y \notin u^{-1} L$. Therefore $u^{-1} L = x^{-1} L$ and $u \in Q(x)$, proving (2.4).

Since L is recognisable, Proposition III.4.17 shows that L has only finitely many quotients on the right and on the left. It follows that only finitely many unions and intersections occur in Formulas (2.3) and (2.4). Furthermore, Lemma 2.6 shows that all the quotients of L are of level n. Consequently, the subsets of the form $Q(x)$ and their quotients are of level n. Formula 2.3 now gives an expression of $S_p(\varphi)$ as a finite union of products of the form $X_0 b_1 X_1 \cdots b_r X_r$ where $b_j \in C_{I_j}$ and the X_0, ..., X_r are level n subsets of D^*. By Formula 2.2, $S_{p-r}(\exists x_{i_1} \cdots \exists x_{i_r} \varphi)$ is thus a finite union of subsets of the form $\pi(X_0 b_1 X_1 b_2 \cdots b_r X_r)$. Therefore, proving that $S_{p-r}(\exists x_{i_1} \cdots \exists x_{i_k} \varphi)$ is of level $n + 1/2$, amounts to proving that if X is a subset of B_p^* of level n contained in D^*, then $\pi(X)$ is a subset of B_{p-k}^* of level n.

Let $\delta : B_{p-k} \to B_p$ be the function defined by

$$\delta(b_0, \ldots, b_{p-k}) = (b_0, 0, \ldots, 0, b_1, \ldots, b_{p-k})$$

Then we have, for each subset X of D^*,

$$\pi(X) = \delta^{-1}(X)$$

Thus, it suffices to verify by induction that the subsets of level n $[n + 1/2]$ are stable under δ^{-1}. For level 0, this follows from the formulas

$$\begin{aligned}
\delta^{-1}(K(\varepsilon_1, \ldots, \varepsilon_{p-k})) &= K(0, \ldots, 0, \varepsilon_{k+1}, \ldots, \varepsilon_p) \\
\delta^{-1}(K(a, i, \varepsilon_1, \ldots, \varepsilon_{p-k})) &= K(a, k+i, 0, \ldots, 0, \varepsilon_{k+1}, \ldots, \varepsilon_p) \\
\delta^{-1}(K_=(i, j, \varepsilon_1, \ldots, \varepsilon_{p-k})) &= K_=(k+i, k+j, 0, \ldots, 0, \varepsilon_{k+1}, \ldots, \varepsilon_p) \\
\delta^{-1}(K_=(i, j, \varepsilon_1, \ldots, \varepsilon_{p-k})) &= K_<(k+i, k+j, 0, \ldots, 0, \varepsilon_{k+1}, \ldots, \varepsilon_p)
\end{aligned}$$

and from the fact that δ^{-1} commutes with the Boolean operations. This latter property also permits one to pass from level $n+1/2$ to level $n+1$ in the induction step. Passing from level n to level $n + 1/2$, requires the formula

$$\delta^{-1}(X_0 b_1 X_1 \cdots b_k X_k) = \bigcup_{\substack{c_1 \in \delta^{-1}(b_1) \\ \vdots \\ c_k \in \delta^{-1}(b_k)}} \delta^{-1}(X_0) c_1 \delta^{-1}(X_1) \cdots c_k \delta^{-1}(X_k)$$

which follows from the fact that δ is length preserving. □

We can now conclude the second part of the proof of Theorem 2.4. Since the only variable-free formulas of $\mathcal{B}\Sigma_0$ are **true** and **false**, the only subsets of A^* definable by a formula of $\mathcal{B}\Sigma_0[<]$ are $\emptyset$ and A^*, which are indeed of level 0.

By Lemma 2.7, for each formula φ of $\mathcal{B}\Sigma_0[<]$, $L_p(\varphi)$ is a subset of B_p^* of level 0. Suppose by induction that, for each formula φ of $\mathcal{B}\Sigma_n[<]$, $L_p(\varphi)$ is a subset of B_p^* of level n. Consider the set Δ of formulas φ such that, for each $p \geqslant 0$, $L_p(\varphi)$ is a subset of B_p^* of level $n + 1/2$. The set Δ contains $\mathcal{B}\Sigma_n[<]$ by the induction hypothesis and, by Proposition IX.3.8, it is closed under finite disjunction and finite conjunction. Finally, Lemma 2.8 shows that if $\varphi \in \Delta$ and if x_1, ..., x_k are free variables of φ, then $\exists x_1 \cdots \exists x_k \varphi \in \Delta$. Therefore Δ contains $\Sigma_{n+1}[<]$ and each formula of $\Sigma_{n+1}[<]$ defines a subset B_p^* of level $n + 1/2$.

To conclude, suppose by induction that for each formula φ of $\Sigma_{n+1}[<]$, $L_p(\varphi)$ is a subset of B_p^* of level $n + 1/2$. The proof of Proposition IX.3.7 shows that

K_p is of level 1 and thus Proposition IX.3.8 shows that for each formula φ of $\mathcal{B}\Sigma_{n+1}[<]$, $L_p(\varphi)$ is a subset of B_p^* of level $n+1$. □

Annex A

A transformation semigroup

We compute in this section the structure of a 4-generator transformation semigroup S. We compute successively the list of its elements, a presentation for it, its right and left Cayley graphs, its $\mathcal{D}$-class structure and the lattice of its idempotents.

$*\ 1$	1	2	3	4
a	2	3	4	0
b	3	1	4	0
c	2	1	4	3
a^2	3	4	0	0
ab	1	4	0	0
ac	1	4	3	0
ba	4	2	0	0
b^2	4	3	0	0
bc	4	2	3	0
ca	3	2	0	4
cb	1	3	0	4
a^3	4	0	0	0
aba	2	0	0	0
ab^2	3	0	0	0
abc	2	3	0	0
aca	2	0	4	0
acb	3	0	4	0
ba^2	0	3	0	0
bab	0	1	0	0
bac	3	1	0	0
b^2a	0	4	0	0
bca	0	3	4	0
bcb	0	1	4	0
cab	4	1	0	0

	1	2	3	4
cac	4	1	0	3
cba	2	4	0	0
cbc	2	4	0	3
$*\ a^4$	0	0	0	0
$*\ bab$	1	0	0	0
$*\ cac$	1	0	3	0
$acbc$	4	0	3	0
$*\ baba$	0	2	0	0
$bcac$	0	4	3	0
$*\ bcbc$	0	2	3	0
$cabc$	3	2	0	0
$*\ caca$	0	2	0	4
$cacb$	0	3	0	4
$cbac$	1	3	0	0
$cbca$	3	0	0	4
$*\ cbcb$	1	0	0	4
$acbca$	0	0	4	0
$cacac$	0	1	0	3
$cacbc$	0	4	0	3
$cbcac$	4	0	0	3
$cbcbc$	2	0	0	3
$*\ acbcac$	0	0	3	0
$*\ cacbca$	0	0	0	4
$cacbcac$	0	0	0	3

Relations :

$$
\begin{array}{cccc}
c^2 = 1 & a^2b = a^3 & a^2c = b^2 & b^3 = b^2a \\
b^2c = a^2 & ca^2 = b^2 & cb^2 = a^2 & a^4 = 0 \\
aba^2 = ab^2 & abac = abab & ab^2a = a^3 & abca = a^2 \\
abcb = ab & acab = abab & acba = a^3 & ba^3 = b^2a \\
bab^2 = ba^2 & babc = baba & baca = ba & bacb = b^2 \\
b^2a^2 = 0 & b^2ab = 0 & b^2ac = ba^2 & bcab = b^2a \\
bcba = baba & caba = baba & cab^2 = ba^2 & cba^2 = ab^2 \\
cbab = abab & ababa = aba & acaca = aca & acacb = acb \\
acbcb = acbca & babab = bab & bcaca = acbca & bcacb = acbca \\
bcbca = bca & bcbcb = bcb & &
\end{array}
$$

The idempotents of S are 1, $a^4 = 0$ and

$$
\begin{array}{cccc}
e_1 = acac & e_2 = cbcb & e_3 = caca & e_4 = bcbc \\
e_5 = abab & e_6 = cacbca & e_7 = baba & e_8 = acbcac
\end{array}
$$

Recall that the natural order on idempotents is defined as follows: $e \leqslant f$ if and only if $ef = e = fe$ (or, equivalently, if $fef = e$).

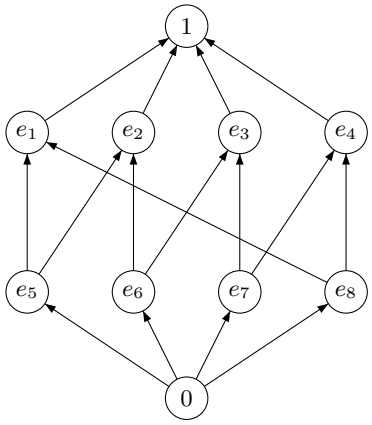


Figure A.1. The lattice of idempotents.

The $\mathcal{D}$-class structure of S is represented in Figure A.2. Note that the nonregular elements bac and $cbac$ have the same kernel and the same range but are not $\mathcal{H}$-equivalent.

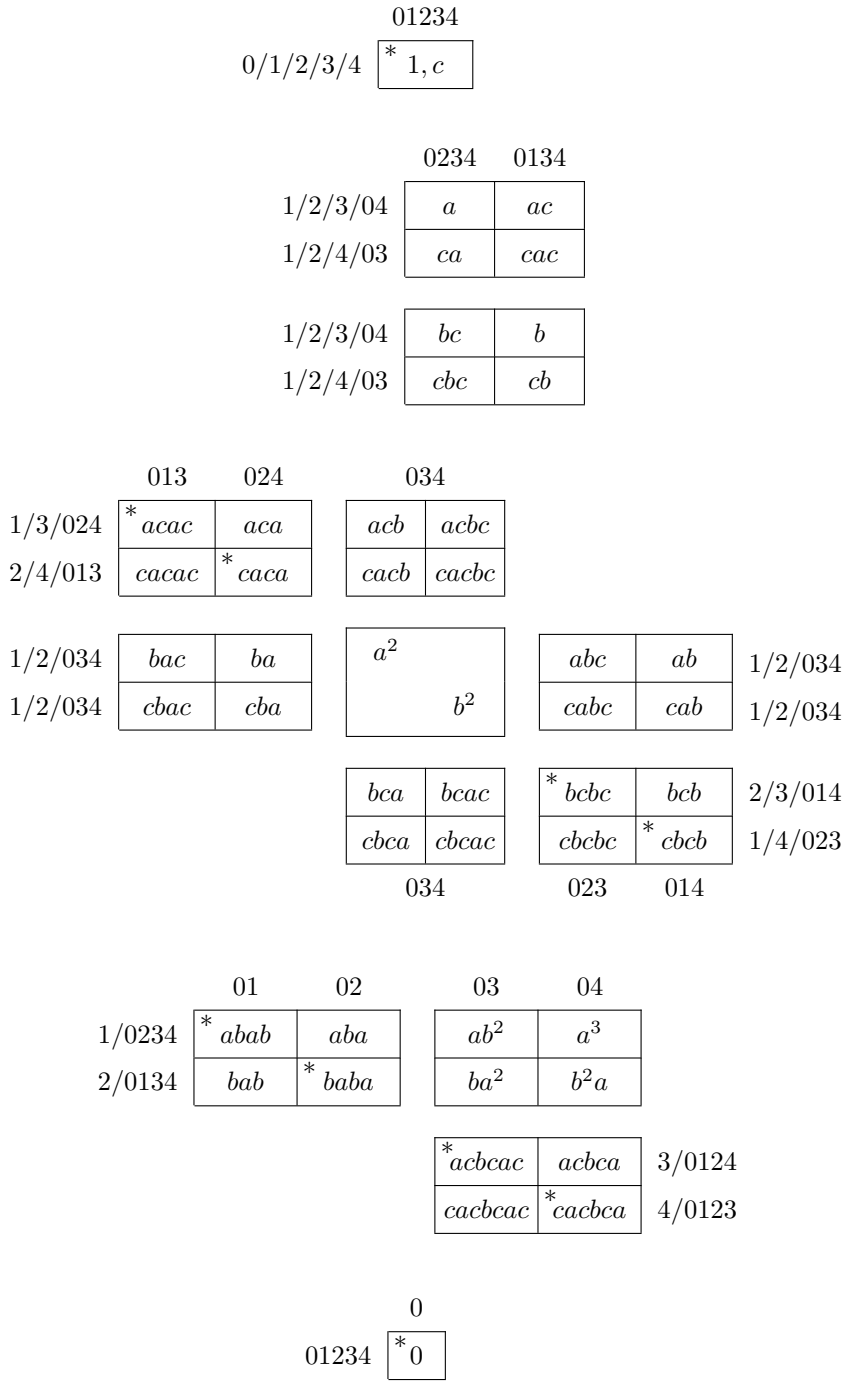


Figure A.2. The $\mathcal{D}$-class structure of S.

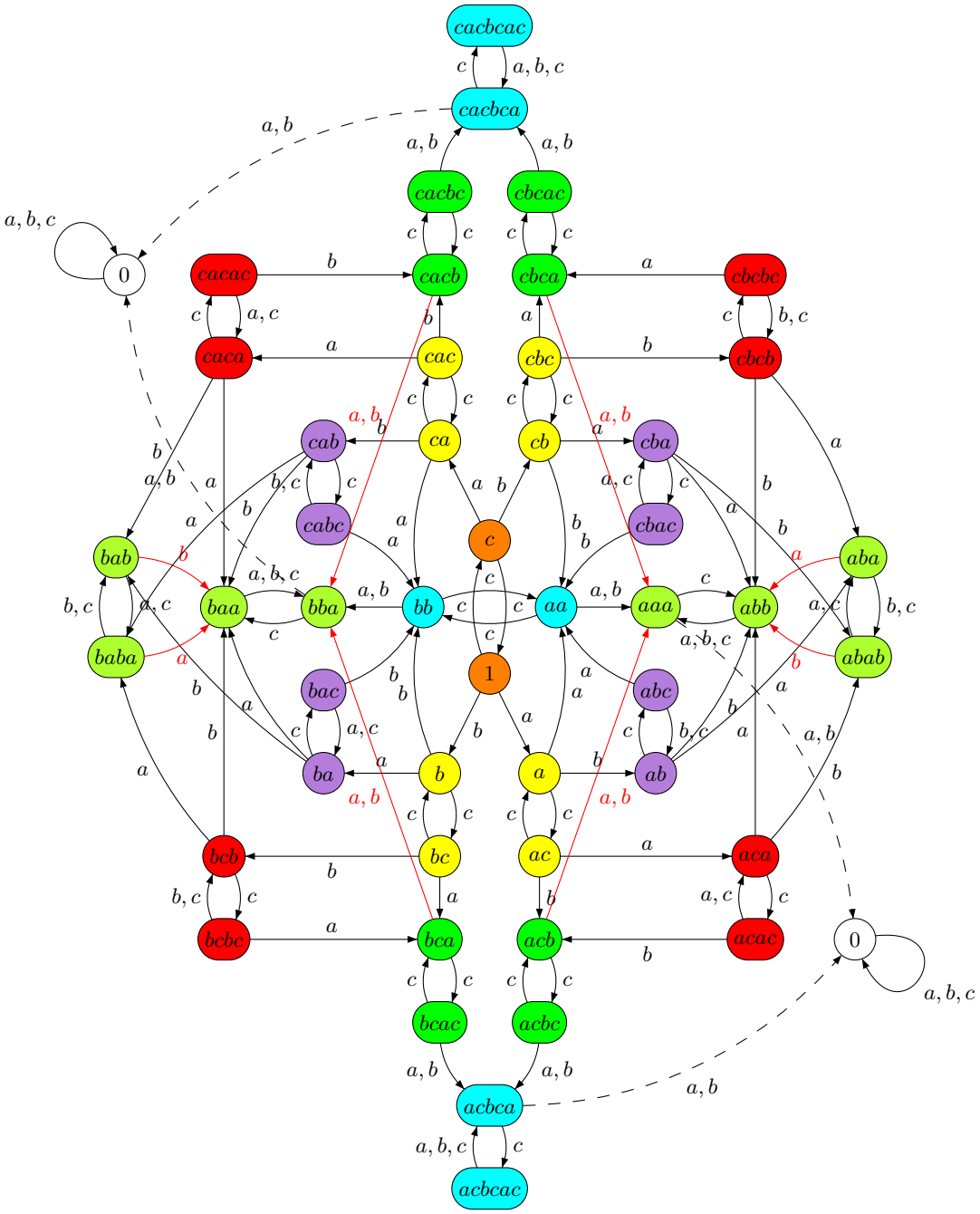


Figure A.3. The right Cayley graph of S. To avoid unesthetic crossing lines, the zero is represented twice in this diagram.

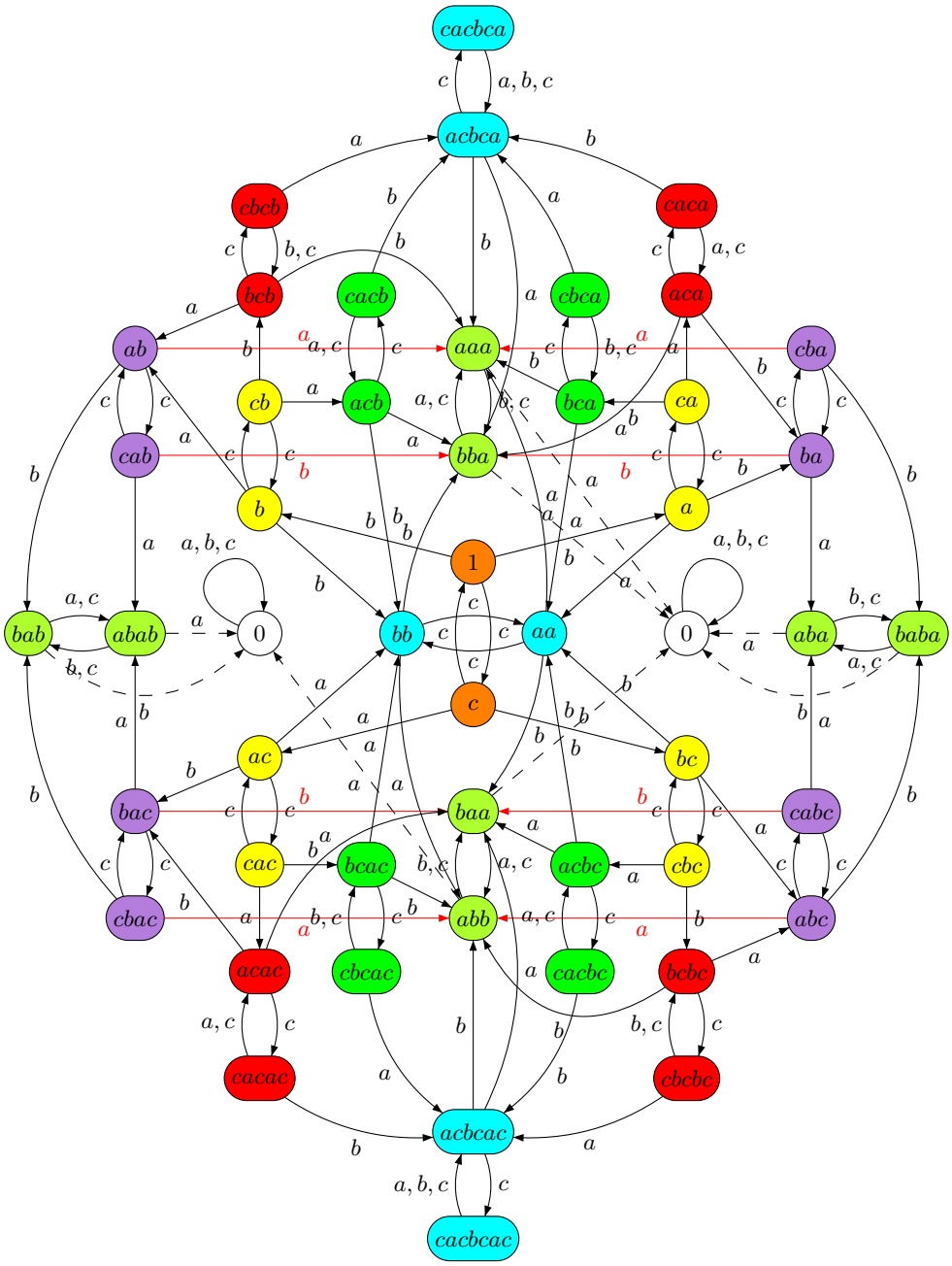


Figure A.4. The left Cayley graph of S. To avoid unesthetic crossing lines, the zero is represented twice in this diagram.

Bibliography

[1] *American Mathematical Society Translations. Series 2, Vol. 59: Twelve papers on logic and algebra*, American Mathematical Society, Providence, R.I., 1966.

[2] J. Almeida, Residually finite congruences and quasi-regular subsets in uniform algebras, *Portugal. Math.* **46**,3 (1989), 313–328. 188

[3] J. Almeida, Implicit operations on finite $\mathcal{J}$-trivial semigroups and a conjecture of I. Simon, *J. Pure Appl. Algebra* **69**,3 (1991), 205–218. 149

[4] J. Almeida, *Finite semigroups and universal algebra*, World Scientific Publishing Co. Inc., River Edge, NJ, 1994. Translated from the 1992 Portuguese original and revised by the author. 6, 10, 188, 207

[5] J. Almeida, Profinite semigroups and applications, in *Structural theory of automata, semigroups, and universal algebra*, Dordrecht, 2005, pp. 1–45, *NATO Sci. Ser. II Math. Phys. Chem.* vol. 207, Springer. Notes taken by Alfredo Costa. 188

[6] J. Almeida and M. V. Volkov, Profinite identities for finite semigroups whose subgroups belong to a given pseudovariety, *J. Algebra Appl.* **2**,2 (2003), 137–163. 214

[7] J. Almeida and P. Weil, Relatively free profinite monoids: an introduction and examples, in *NATO Advanced Study Institute Semigroups, Formal Languages and Groups*, J. Fountain (ed.), vol. 466, pp. 73–117, Kluwer Academic Publishers, 1995.

[8] M. Arfi, Polynomial operations on rational languages, in *STACS 87 (Passau, 1987)*, pp. 198–206, *Lecture Notes in Comput. Sci.* vol. 247, Springer, Berlin, 1987. 297, 298

[9] M. Arfi, Opérations polynomiales et hiérarchies de concaténation, *Theoret. Comput. Sci.* **91**,1 (1991), 71–84. 297, 298

[10] B. Banaschewski, The Birkhoff theorem for varieties of finite algebras, *Algebra Universalis* **17**,3 (1983), 360–368. 189

[11] M.-P. Béal, O. Carton and C. Reutenauer, Cyclic languages and strongly cyclic languages, in *STACS 96 (Grenoble, 1996)*, Berlin, 1996, pp. 49–59, *Lect. Notes Comp. Sci.* vol. 1046, Springer. 247

[12] D. Beauquier and J.-É. Pin, Languages and scanners, *Theoret. Comput. Sci.* **84** (1991), 3–21. 155

[13] J. Berstel, *Transductions and context-free languages*, Teubner, 1979. 97

[14] J. Berstel, D. Perrin and C. Reutenauer, *Codes and Automata*, *Encyclopedia of Mathematics and its Applications* vol. 129, Cambridge University Press, 2009. 634 pages.

[15] G. Birkhoff, On the structure of abstract algebras, *Proc. Cambridge Phil. Soc.* **31** (1935), 433–454. 4

[16] G. Birkhoff, Moore-Smith convergence in general topology, *Annals of Mathematics* **38** (1937), 39–56. 188

[17] B. Borchert, D. Kuske and F. Stephan, On existentially first-order definable languages and their relation to NP, in *Automata, languages and programming (Aalborg, 1998)*, pp. 17–28, *Lecture Notes in Comput. Sci.* vol. 1443, Springer, Berlin, 1998. 297

[18] M. J. Branco and J.-É. Pin, Equations for the polynomial closure, in *ICALP 2009, Part II*, S. Albers and W. Thomas (eds.), Berlin, 2009, pp. 115–126, *Lect. Notes Comp. Sci.* vol. 5556, Springer. 256

[19] J. Brzozowski, K. Culik and A. Gabrielan, Classification of noncounting events, *J. Comput. Syst. Sci.* **5** (1971), 41–53.

[20] J. A. Brzozowski, Hierarchies of Aperiodic Languages, *Theoret. Informatics Appl.* **10**,2 (1976), 33–49.

[21] J. A. Brzozowski and F. E. Fich, Languages of R-Trivial Monoids, *J. Comput. Syst. Sci.* **20**,1 (1980), 32–49.

[22] J. A. Brzozowski and R. Knast, The dot-depth hierarchy of star-free languages is infinite, *J. Comput. System Sci.* **16**,1 (1978), 37–55. 298

[23] J. A. Brzozowski and I. Simon, Characterizations of locally testable events, *Discrete Math.* **4** (1973), 243–271. 3, 154, 155

[24] J. R. Büchi, Weak second-order arithmetic and finite automata, *Z. Math. Logik und Grundl. Math.* **6** (1960), 66–92. 157

[25] J. R. Büchi, On a decision method in restricted second order arithmetic, in *Logic, Methodology and Philosophy of Science (Proc. 1960 Internat. Congr .)*, Stanford, Calif., 1962, pp. 1–11, Stanford Univ. Press.

[26] J. Carroll and D. Long, *Theory of finite automata with an introduction to formal languages*, Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1989. 6

[27] O. Carton, *Langages formels, calculabilité et complexité*, Vuibert, 2008. 5, 74

[28] O. Carton, J.-E. Pin and X. Soler-Escrivà, Languages Recognized by Finite Supersoluble Groups, *Journal of Automata, Languages and Combinatorics* **14**,2 (2009), 149–161.

[29] J. Chalopin and H. Leung, On factorization forests of finite height, *Theoret. Comput. Sci.* **310**,1-3 (2004), 489–499. 34

[30] A. H. Clifford and G. B. Preston, *The Algebraic Theory of Semigroups*, vol. 1, Amer. Math. Soc., 1961. 6, 128

[31] A. H. Clifford and G. B. Preston, *The Algebraic Theory of Semigroups*, vol. 2, Amer. Math. Soc., 1967. 6, 128

[32] J. Cohen, D. Perrin and J.-E. Pin, On the expressive power of temporal logic, *J. Comput. System Sci.* **46**,3 (1993), 271–294. 281

[33] R. S. Cohen and J. A. Brzozowski, On star-free events, in *Proc. Hawaii Int. Conf. on System Science*, B. K. Kinariwala and F. F. Kuo (eds.), pp. 1–4, University of Hawaii Press, Honolulu, HI, 1968. 138, 281

[34] T. Colcombet, A combinatorial theorem for trees: applications to monadic logic and infinite structures, in *Automata, languages and programming*, Berlin, 2007, pp. 901–912, *Lect. Notes Comp. Sci.* vol. 4596, Springer. 34

[35] T. Colcombet, Factorisation Forests for Infinite Words, in *Fundamentals of Computation Theory, 16th International Symposium, FCT 2007, Budapest, Hungary, August 27-30, 2007, Proceedings*, E. Csuhaj-Varjú and Z. Ésik (eds.), Berlin, 2007, pp. 226–237, *Lect. Notes Comp. Sci.* vol. 4639, Springer. 34

[36] J. H. Conway, *Regular Algebra and Finite Machines*, Chapman and Hall, London, 1971. 42

[37] D. M. Davenport, On power commutative semigroups, *Semigroup Forum* **44**,1 (1992), 9–20.

[38] A. De Luca and S. Varricchio, *Finiteness and Regularity in Semigroups and Formal Languages*, Springer-Verlag, 1999.

[39] V. Diekert and P. Gastin, Pure future local temporal logics are expressively complete for Mazurkiewicz traces, *Inform. and Comput.* **204**,11 (2006), 1597–1619. Conference version in LATIN 2004, LNCS 2976, 170–182, 2004.

[40] V. Diekert and M. Kufleitner, A survey on the local divisor technique, *Theoret. Comput. Sci.* **610**,part A (2016), 13–23. 138

[41] S. Eilenberg, *Automata, languages, and machines. Vol. A*, Academic Press [A subsidiary of Harcourt Brace Jovanovich, Publishers], New York, 1974. Pure and Applied Mathematics, Vol. 58. 97

[42] S. Eilenberg, *Automata, languages, and machines. Vol. B*, Academic Press [Harcourt Brace Jovanovich Publishers], New York, 1976. With two chapters ("Depth decomposition theorem" and "Complexity of semigroups and morphisms") by Bret Tilson, Pure and Applied Mathematics, Vol. 59. 4, 6, 75, 189, 207, 226, 230, 235, 295, 297

[43] C. C. Elgot, Decision problems of finite automata design and related arithmetics, *Trans. Amer. Math. Soc.* **98** (1961), 21–51. 157

[44] Z. Ésik, Extended temporal logic on finite words and wreath products of monoids with distinguished generators, in *DLT 2002, Kyoto, Japan*, M. E. A. Ito (ed.), Berlin, 2002, pp. 43–58, *Lect. Notes Comp. Sci.* n° 2450, Springer. 230

[45] Z. Ésik and M. Ito, Temporal logic with cyclic counting and the degree of aperiodicity of finite automata, *Acta Cybernetica* **16** (2003), 1–28. 4

[46] Z. Ésik and K. G. Larsen, Regular languages definable by Lindström quantifiers, *Theoret. Informatics Appl.* **37** (2003), 179–241.

[47] M. Gehrke, S. Grigorieff and J.-É. Pin, Duality and equational theory of regular languages, in *ICALP 2008, Part II*, L. A. et al. (ed.), Berlin, 2008, pp. 246–257, *Lect. Notes Comp. Sci.* vol. 5126, Springer. 4, 222

[48] S. Ginsburg and E. H. Spanier, Bounded ALGOL-like languages, *Trans. Amer. Math. Soc.* **113** (1964), 333–368.

[49] S. Ginsburg and E. H. Spanier, Bounded regular sets, *Proc. Amer. Math. Soc.* **17** (1966), 1043–1049. 220

[50] S. Ginsburg and E. H. Spanier, Semigroups, Presburger formulas, and languages, *Pacific J. Math.* **16** (1966), 285–296.

[51] A. Ginzburg, *Algebraic theory of automata*, Academic Press, New York, 1968. 6

[52] C. Glasser and H. Schmitz, Languages of dot-depth 3/2, in *STACS 2000 (Lille)*, pp. 555–566, *Lecture Notes in Comput. Sci.* vol. 1770, Springer, Berlin, 2000.

[53] R. Graham, On finite 0-simple semigroups and graph theory, *Math. Syst. Theory* **2** (1968), 325–339.

[54] J. A. Green, On the structure of semigroups, *Ann. of Math. (2)* **54** (1951), 163–172. 99

[55] P. A. Grillet, *Semigroups, An Introduction to the Structure Theory*, Marcel Dekker, Inc., New York, 1995. 6, 128

[56] M. Hall, A Topology for free groups and related groups, *Annals of Mathematics* **52** (1950), 127–139. 188

[57] K. Henckell and J.-E. Pin, Ordered monoids and $\mathcal{J}$-trivial monoids, in *Algorithmic problems in groups and semigroups (Lincoln, NE, 1998)*, pp. 121–137, *Trends Math.*, Birkhäuser Boston, Boston, MA, 2000. 149

[58] K. Henckell and J. Rhodes, The theorem of Knast, the $PG = BG$ and type-II conjectures, in *Monoids and semigroups with applications (Berkeley, CA, 1989)*, pp. 453–463, World Sci. Publ., River Edge, NJ, 1991.

[59] J. B. Hickey, Semigroups under a sandwich operation, *Proc. Edinburgh Math. Soc. (2)* **26**,3 (1983), 371–382. 35

[60] Higgins, *Techniques of Semigroup Theory*, World Scientific, 1999. 6, 128

[61] P. M. Higgins, A proof of Simon's theorem on piecewise testable languages, *Theoret. Comput. Sci.* **178**,1-2 (1997), 257–264.

[62] P. M. Higgins, A new proof of Schützenberger's theorem, *Internat. J. Algebra Comput.* **10**,2 (2000), 217–220.

[63] J. Hopcroft, R. Motwani and J. D. Ullman, *Introduction to automata theory, languages, and computation*, Addison-Wesley, Boston, ed. 2nd, 2001. 6

[64] J. Hopcroft and J. D. Ullman, *Introduction to Automata Theory, Languages and Computation*, Addison Wesley, 1979. 58, 74

[65] J. M. Howie, *An Introduction to Semigroup Theory*, Academic Press, 1976.

[66] J. M. Howie, *Automata and Languages*, Clarendon Press, 1991.

[67] J. M. Howie, *Fundamentals of semigroup theory*, *London Mathematical Society Monographs. New Series* vol. 12, The Clarendon Press Oxford University Press, New York, 1995. Oxford Science Publications.

[68] L. (http://math.stackexchange.com/users/17760/lubin), Reference request for tricky problem in elementary group theory. Mathematics Stack Exchange. URL:http://math.stackexchange.com/q/253514 (version: 2016-02-29). 127

[69] S. C. Kleene, Representation of events in nerve nets and finite automata, in *Automata studies*, Princeton, N. J., 1956, pp. 3–41, Princeton University Press. Annals of mathematics studies, no. 34. 3

[70] O. Klíma, Piecewise testable languages via combinatorics on words, *Discrete Math.* **311**,20 (2011), 2124–2127. 149

[71] R. Knast, A semigroup characterization of dot-depth one languages, *RAIRO Inform. Théor.* **17**,4 (1983), 321–330.

[72] R. Knast, Some theorems on graph congruences, *RAIRO Inform. Théor.* **17**,4 (1983), 331–342.

[73] D. Krob, Complete sets of B-rational identities, *Theoret. Comp. Sci.* **89** (1991), 207–343. 42

[74] J. B. Kruskal, The theory of well-quasi-ordering: A frequently discovered concept, *J. Combinatorial Theory Ser. A* **13** (1972), 297–305.

[75] M. Kufleitner, The Height of Factorization Forests, in *Mathematical Foundations of Computer Science 2008, 33rd International Symposium, MFCS 2008, Torun, Poland, August 25-29, 2008, Proceedings*, E. Ochmanski and J. Tyszkiewicz (eds.), Berlin, 2008, pp. 443–454, *Lect. Notes Comp. Sci.* vol. 5162, Springer. 34

[76] M. Kufleitner and A. Lauser, Lattices of Logical Fragments over Words, in *ICALP 2012, Part II*, Berlin, Heidelberg, 2012, pp. 275–286, *Lect. Notes Comp. Sci.* vol. 7392, Springer. 171

[77] M. Kunc, Equational description of pseudovarieties of homomorphisms, *Theoretical Informatics and Applications* **37** (2003), 243–254.

[78] G. Lallement, *Semigroups and Combinatorial Applications*, Wiley and Sons, 1979. 6, 128

[79] M. V. Lawson, *Finite Automata*, CRC Press, 2003.

[80] E. Le Rest and M. Le Rest, Sur le calcul du monoïde syntaxique d'un sous monoïde finiment engendré, *Semigroup Forum* **21**,2-3 (1980), 173–185.

[81] M. Lothaire, *Combinatorics on Words*, *Encyclopedia of Mathematics and its Applications* vol. 17, Cambridge University Press, 1983. 7

[82] M. Lothaire, *Combinatorics on words*, *Cambridge Mathematical Library*, Cambridge University Press, Cambridge, 1997. With a foreword by Roger Lyndon and a preface by Dominique Perrin, Corrected reprint of the 1983 original, with a new preface by Perrin.

[83] S. Margolis and J.-É. Pin, Products of group languages, in *FCT*, Berlin, 1985, pp. 285–299, *Lect. Notes Comp. Sci.* n° 199, Springer. 297

[84] Mascle, Torsion Matrix Semigroups and Recognizable Transductions, in *Automata, Languages and Programming*, Kott (ed.), Berlin, 1986, pp. 244–253, *Lect. Notes Comp. Sci.*, Springer.

[85] R. McNaughton, Algebraic decision procedures for local testability, *Math. Systems Theory* **8**,1 (1974), 60–76. 3, 154, 155

[86] R. McNaughton and S. Papert, *Counter-free automata*, The M.I.T. Press, Cambridge, Mass.-London, 1971. With an appendix by William Henneman, M.I.T. Research Monograph, No. 65. 3, 157

[87] R. McNaughton and H. Yamada, Regular Expressions and State Graphs for Automata, *IRE Trans. Electronic Computers* **9**,1 (1960), 39–47.

[88] A. R. Meyer, A note on star-free events, *J. Assoc. Comput. Mach.* **16** (1969), 220–225. 138, 281

[89] M. Morse and G. A. Hedlund, Unending chess, symbolic dynamics and a problem in semigroups, *Duke Math. J.* **11** (1944), 1–7.

[90] G. Păun and A. Salomaa, Thin and slender languages, *Discrete Appl. Math.* **61**,3 (1995), 257–270. 220

[91] M. Perles, M. Rabin and E. Shamir, The theory of definite automata, *IEEE Trans. Electron. Comput.* **12** (1963), 233–243. 3

[92] D. Perrin, Finite automata, in *Handbook of theoretical computer science, Vol. B*, pp. 1–57, Elsevier, Amsterdam, 1990.

[93] D. Perrin and J.-É. Pin, *Infinite Words*, *Pure and Applied Mathematics* vol. 141, Elsevier, 2004. ISBN 0-12-532111-2. 7

[94] M. Petrich, *Inverse Semigroups*, Wiley and Sons, 1984.

[95] S. Piccard, Sur les fonctions définies dans les ensembles finis quelconques, *Fundam. Math.* **24** (1935), 298–301.

[96] J.-É. Pin, Propriétés syntactiques du produit non ambigu, in *7th ICALP*, Berlin, 1980, pp. 483–499, *Lect. Notes Comp. Sci.* n° 85, Springer.

[97] J.-E. Pin, Hiérarchies de concaténation, *RAIRO Inform. Théor.* **18**,1 (1984), 23–46.

[98] J.-É. Pin, *Varieties of formal languages*, Plenum Publishing Corp., New York, 1986. With a preface by M.-P. Schützenberger, Translated from the French by A. Howie. 7, 207

[99] J.-É. Pin, A property of the Schützenberger product, *Semigroup Forum* **35** (1987), 53–62. 265

[100] J.-E. Pin, Finite semigroups and recognizable languages: an introduction, in *Semigroups, formal languages and groups (York, 1993)*, pp. 1–32, Kluwer Acad. Publ., Dordrecht, 1995.

[101] J.-E. Pin, A variety theorem without complementation, *Russian Mathematics (Iz. VUZ)* **39** (1995), 80–90. 4, 207, 297

[102] J.-E. Pin, The expressive power of existential first order sentences of Büchi's sequential calculus, in *Automata, languages and programming (Paderborn, 1996)*, Berlin, 1996, pp. 300–311, *Lecture Notes in Comput. Sci.* vol. 1099, Springer.

[103] J.-É. Pin, Logic, Semigroups and Automata on Words, *Annals of Mathematics and Artificial Intelligence* **16** (1996), 343–384.

[104] J.-É. Pin, Polynomial closure of group languages and open sets of the Hall topology, *Theoret. Comput. Sci.* **169** (1996), 185–200. Journal version of the article of ICALP 1994. 297

[105] J.-É. Pin, Syntactic semigroups, in *Handbook of formal languages*, G. Rozenberg and A. Salomaa (eds.), vol. 1, ch. 10, pp. 679–746, Springer Verlag, 1997.

[106] J.-É. Pin, Bridges for concatenation hierarchies, in *25th ICALP*, Berlin, 1998, pp. 431–442, *Lect. Notes Comp. Sci.* n° 1443, Springer.

[107] J.-É. Pin, Logic On Words, in *Current Trends in Theoretical Computer Science, Entering the 21st Century*, G. R. G. Păun and A. Salomaa (eds.), pp. 254–273, Word Scientific, 2001.

[108] J.-É. Pin, Algebraic tools for the concatenation product, *Theoret. Comput. Sci.* **292** (2003), 317–342. 137, 265

[109] J.-É. Pin, The expressive power of existential first order sentences of Büchi's sequential calculus, *Discrete Mathematics* **291** (2005), 155–174.

[110] J.-É. Pin, Profinite methods in automata theory, in *26th International Symposium on Theoretical Aspects of Computer Science (STACS 2009)*, S. Albers (ed.), Dagstuhl, Germany, 2009, pp. 31–50, Internationales Begegnungs- Und Forschungszentrum für Informatik (IBFI), Schloss Dagstuhl, Germany.

[111] J.-É. Pin, A. Pinguet and P. Weil, Ordered categories and ordered semigroups, *Communications in Algebra* **30** (2002), 5651–5675.

[112] J.-É. Pin and H. Straubing, Monoids of upper triangular matrices, in *Semigroups (Szeged, 1981)*, Amsterdam, 1985, pp. 259–272, *Colloq. Math. Soc. János Bolyai* vol. 39, North-Holland.

[113] J.-É. Pin and H. Straubing, Some results on $\mathcal{C}$-varieties, *Theoret. Informatics Appl.* **39** (2005), 239–262.

[114] J.-É. Pin, H. Straubing and D. Thérien, Some results on the generalized star-height problem, *Information and Computation* **101** (1992), 219–250.

[115] J.-É. Pin and P. Weil, Polynomial closure and unambiguous product, in *22th ICALP*, Berlin, 1995, pp. 348–359, *Lect. Notes Comp. Sci.* n° 944, Springer. 298

[116] J.-É. Pin and P. Weil, Profinite semigroups, Mal'cev products and identities, *J. of Algebra* **182** (1996), 604–626.

[117] J.-É. Pin and P. Weil, A Reiterman theorem for pseudovarieties of finite first-order structures, *Algebra Universalis* **35** (1996), 577–595.

[118] J.-É. Pin and P. Weil, Polynomial closure and unambiguous product, *Theory Comput. Systems* **30** (1997), 1–39. 265, 297

[119] J.-É. Pin and P. Weil, A conjecture on the concatenation product, *ITA* **35** (2001), 597–618.

[120] J.-É. Pin and P. Weil, Semidirect products of ordered semigroups, *Communications in Algebra* **30** (2002), 149–169.

[121] J.-É. Pin and P. Weil, The wreath product principle for ordered semigroups, *Communications in Algebra* **30** (2002), 5677–5713.

[122] N. Pippenger, Regular languages and Stone duality, *Theory Comput. Syst.* **30**,2 (1997), 121–134.

[123] T. Place, Separating Regular Languages with Two Quantifier Alternations, in *30th Annual ACM/IEEE Symposium on Logic in Computer Science (LICS 2015)*, pp. 202–213, IEEE, 2015. 298

[124] T. Place, L. Van Rooijen and M. Zeitoun, Separating regular languages by locally testable and locally threshold testable languages, in *33nd International Conference on Foundations of Software Technology and Theoretical Computer Science*, pp. 363–375, *LIPIcs. Leibniz Int. Proc. Inform.* vol. 24, Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2013.

[125] T. Place, L. Van Rooijen and M. Zeitoun, Separating regular languages by piecewise testable and unambiguous languages, in *Mathematical foundations of computer science 2013*, pp. 729–740, *Lect. Notes Comp. Sci.* vol. 8087, Springer, Heidelberg, 2013.

[126] T. Place, L. Van Rooijen and M. Zeitoun, On separation by locally testable and locally threshold testable languages, *Log. Methods Comput. Sci.* **10**,3 (2014), 3:24, 28.

[127] T. Place and M. Zeitoun, Going higher in the first-order quantifier alternation hierarchy on words, in *Automata, languages, and programming. Part II*, pp. 342–353, *Lect. Notes Comp. Sci.* vol. 8573, Springer, Heidelberg, 2014. 298

[128] T. Place and M. Zeitoun, Separating regular languages with first-order logic, in *Joint Meeting of the Twenty-Third EACSL Annual Conference on Computer Science Logic (CSL) and the Twenty-Ninth Annual ACM/IEEE Symposium on Logic in Computer Science (LICS), CSL-LICS '14, Vienna, Austria, July 14 - 18, 2014*, T. A. Henzinger and D. Miller (eds.), pp. 75:1–75:10, ACM, 2014.

[129] T. Place and M. Zeitoun, Separation and the Successor Relation, in *32nd International Symposium on Theoretical Aspects of Computer Science, STACS 2015, March 4-7, 2015, Garching, Germany*, E. W. Mayr and N. Ollinger (eds.), pp. 662–675, *LIPIcs* vol. 30, Schloss Dagstuhl - Leibniz-Zentrum Fuer Informatik, 2015.

[130] L. Polák, Syntactic semiring of a language, in *MFCS 2001*, Berlin, 2001, pp. 611–620, *Lect. Notes Comp. Sci.* vol. 2136, Springer.

[131] M. O. Rabin and D. Scott, Finite automata and their decision problems, Rap. Tech., IBM J. Res. and Develop., 1959. "Reprinted in *Sequential Machines, E. F. Moore (ed.)*, Addison-Wesley, Reading, Massachussetts, (1964), 63–91.". 3, 77

[132] N. R. Reilly and S. Zhang, Decomposition of the lattice of pseudovarieties of finite semigroups induced by bands, *Algebra Universalis* **44**,3-4 (2000), 217–239. 214

[133] J. Reiterman, The Birkhoff theorem for finite algebras, *Algebra Universalis* **14**,1 (1982), 1–10. 4, 189

[134] C. Reutenauer, Une topologie du monoïde libre, *Semigroup Forum* **18**,1 (1979), 33–49. 188

[135] C. Reutenauer, Sur mon article: "Une topologie du monoïde libre" [Semigroup Forum **18** (1979), no. 1, 33–49; MR 80j:20075], *Semigroup Forum* **22**,1 (1981), 93–95.

[136] J. Rhodes and B. Steinberg, *The q-theory of finite semigroups*, *Springer Monographs in Mathematics*, Springer, New York, 2009. 128

[137] J. Sakarovitch, *Éléments de théorie des automates*, Vuibert, 2003.

[138] J. Sakarovitch, *Elements of automata theory*, Cambridge University Press, Cambridge, 2009. Translated from the 2003 French original by Reuben Thomas. 5, 74, 97

[139] A. Salomaa, *Computation and automata*, *Encyclopedia of Mathematics and its Applications* vol. 25, Cambridge University Press, Cambridge, 1985. With a foreword by Grzegorz Rozenberg. 6

[140] M. P. Schützenberger, Une théorie algébrique du codage, *Séminaire Dubreil. Algèbre et théorie des nombres* **9** (1955-1956), 1–24. http://eudml.org/doc/111094. 3, 97

[141] M.-P. Schützenberger, On finite monoids having only trivial subgroups, *Information and Control* **8** (1965), 190–194. 3, 137

[142] M. P. Schützenberger, Sur le produit de concaténation non ambigu, *Semigroup Forum* **13**,1 (1976/77), 47–75.

[143] C. Selmi, Strongly locally testable semigroups with commuting idempotents and related languages, *Theor. Inform. Appl.* **33**,1 (1999), 47–57.

[144] I. Simon, *Hierarchies of Events with Dot-Depth One*, PhD thesis, University of Waterloo, Waterloo, Ontario, Canada, 1972. 5

[145] I. Simon, Piecewise testable events, in *Proc. 2nd GI Conf.*, H. Brackage (ed.), pp. 214–222, *Lecture Notes in Comp. Sci.* vol. 33, Springer Verlag, Berlin, Heidelberg, New York, 1975. 3, 149, 298

[146] I. Simon, Limited Subsets of a Free Monoid, in *Proc. 19th Annual Symposium on Foundations of Computer Science*, Piscataway, N.J., 1978, pp. 143–150, IEEE.

[147] I. Simon, Properties of factorization forests, in *Formal properties of finite automata and applications, Ramatuelle, France, May 23-27, 1988, Proceedings*, Berlin, 1989, pp. 65–72, *Lect. Notes Comp. Sci.* vol. 386, Springer. 34

[148] I. Simon, Factorization forests of finite height, *Theoret. Comput. Sci.* **72**,1 (1990), 65–94. 34

[149] I. Simon, A short proof of the factorization forest theorem, in *Tree automata and languages (Le Touquet, 1990)*, Amsterdam, 1992, pp. 433–438, *Stud. Comput. Sci. Artificial Intelligence* vol. 10, North-Holland. 34

[150] J. Stern, Characterizations of some classes of regular events, *Theoret. Comput. Sci.* **35**,1 (1985), 17–42.

[151] M. Stone, The theory of representations for Boolean algebras, *Trans. Amer. Math. Soc.* **40** (1936), 37–111.

[152] M. H. Stone, Applications of the theory of Boolean rings to general topology, *Trans. Amer. Math. Soc.* **41**,3 (1937), 375–481.

[153] M. H. Stone, The representation of Boolean algebras, *Bull. Amer. Math. Soc.* **44**,12 (1938), 807–816. 175

[154] H. Straubing, Aperiodic homomorphisms and the concatenation product of recognizable sets, *J. Pure Appl. Algebra* **15**,3 (1979), 319–327. 137

[155] H. Straubing, Families of recognizable sets corresponding to certain varieties of finite monoids, *J. Pure Appl. Algebra* **15**,3 (1979), 305–318. 137, 281

[156] H. Straubing, A generalization of the Schützenberger product of finite monoids, *Theoret. Comput. Sci.* **13**,2 (1981), 137–150. 297

[157] H. Straubing, Relational morphisms and operations on recognizable sets, *RAIRO Inf. Theor.* **15** (1981), 149–159. 265, 268

[158] H. Straubing, Finite semigroup varieties of the form $V * D$, *J. Pure Appl. Algebra* **36**,1 (1985), 53–94. 155, 297

[159] H. Straubing, Semigroups and languages of dot-depth two, *Theoret. Comput. Sci.* **58**,1-3 (1988), 361–378. Thirteenth International Colloquium on Automata, Languages and Programming (Rennes, 1986).

[160] H. Straubing, The wreath product and its applications, in *Formal properties of finite automata and applications (Ramatuelle, 1988)*, Berlin, 1989, pp. 15–24, *Lecture Notes in Comput. Sci.* vol. 386, Springer. 281

[161] H. Straubing, *Finite automata, formal logic, and circuit complexity*, Birkhäuser Boston Inc., Boston, MA, 1994.

[162] H. Straubing, On logical descriptions of regular languages, in *LATIN 2002*, Berlin, 2002, pp. 528–538, *Lect. Notes Comp. Sci.* n° 2286, Springer. 4, 230

[163] H. Straubing and D. Thérien, Partially ordered finite monoids and a theorem of I. Simon, *J. Algebra* **119**,2 (1988), 393–399. 149

[164] H. Straubing and P. Weil, On a conjecture concerning dot-depth two languages, *Theoret. Comput. Sci.* **104**,2 (1992), 161–183.

[165] A. Szilard, S. Yu, K. Zhang and J. Shallit, Characterizing regular languages with polynomial densities, in *Mathematical foundations of computer science 1992 (Prague, 1992)*, pp. 494–503, *Lect. Notes Comp. Sci.* vol. 629, Springer, Berlin, 1992. 220

[166] D. Thérien, Classification of finite monoids: the language approach, *Theoret. Comput. Sci.* **14**,2 ang. (1981), 195–208. 297

[167] D. Thérien and A. Weiss, Graph congruences and wreath products, *J. Pure Appl. Algebra* **36**,2 (1985), 205–215. 155

[168] W. Thomas, Classifying regular events in symbolic logic, *J. Comput. System Sci.* **25**,3 (1982), 360–376. 4, 5, 297

[169] A. Thue, Über unendliche Zeichenreihen, *Norske Vid. Selsk. Skr. I Math-Nat. Kl.* **7** (1906), 1–22.

[170] A. Thue, Über die gegenseitige Loge gleicher Teile gewisser Zeichenreihen, *Norske Vid. Selsk. Skr. I Math-Nat. Kl. Chris.* **1** (1912), 1–67.

[171] B. Tilson, On the complexity of finite semigroups, *J. Pure Appl. Algebra* **5** (1974), 187–208.

[172] B. A. Trahtenbrot, Finite automata and the logic of one-place predicates, *Sibirsk. Mat. Zh.* **3** (1962), 103–131. AMS Transl., 59 (1966).

[173] W. Wechler, *Universal algebra for computer scientists*, *EATCS Monographs on Theoretical Computer Science* vol. 25, Springer-Verlag, Berlin, 1992.

[174] P. Weil, Some results on the dot-depth hierarchy, *Semigroup Forum* **46**,3 (1993), 352–370.

[175] P. Weil, Profinite methods in semigroup theory, *Int. J. Alg. Comput.* **12** (2002), 137–178. 188

[176] Th. Wilke, Classifying discrete Temporal Properties, in *STACS'99*, C. Meinel (ed.), Trier, Germany, 1999, pp. 32–46, *Lecture Notes in Comput. Sci.* vol. 1563, Springer. 281

[177] S. Yu, Regular languages, in *Handbook of language theory*, G. Rozenberg and A. Salomaa (eds.), vol. 1, ch. 2, pp. 679–746, Springer, 1997. 216, 219, 220

Index of Notation

Index