Mathematical Grounds of Automata Theory

Jean-Éric Pin

January 30, 2007

Contents

Introduction

Notation. I use Greek letters for functions, lower case letters for elements, capital letters for sets and cursive 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.

I use the term *morphism* for *homomorphism*. Alphabets are usually denoted by A or B and letters by a, b, c, ...

Fundamentals

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.

We denote by $|E|$ 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.

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 can also be viewed as a function[1] from E into $\mathcal{P}(F)$ 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 into F.

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

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

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

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

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

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

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

If Y is a subset of F, we then have

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

Example 1.1 Let τ be the relation from $E = \{1, 2, 3\}$ into $F = \{1, 2, 3, 4\}$ defined by $\tau = \{(1, 1), (1, 2), (2, 1), (2, 3), (2, 4)\}$.

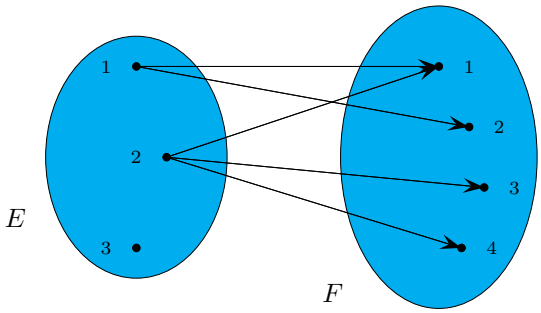


Figure 1.1. The relation τ.

Then $\tau(1) = \{1, 2\}$, $\tau(2) = \{1, 3, 4\}$, $\tau(3) = \emptyset$, $\tau^{-1}(1) = \{1, 2\}$, $\tau^{-1}(2) = \{1\}$, $\tau^{-1}(3) = \{2\}$, $\tau^{-1}(4) = \{2\}$.

Given two relations $\tau_1 : E \to F$ and $\tau_2 : F \to G$, we denote by $\tau_1\tau_2$ or by $\tau_2 \circ \tau_1$ the *composition* of τ_1 and τ_2, which is the relation from E into 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 of domain E is sometimes called a *total function* or a *mapping* from E into 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 $Id_E : E \to E$ defined by $Id_E(x) = x$ 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 $Id_E(x)$ 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 $\varphi \circ \psi = Id_F$.*

Proof. If there exists a mapping ψ with these properties, we have $\varphi(\psi(x)) = x$ for all $x \in E$ 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 into 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 $\varphi \circ \psi = Id_F$. 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 $\psi \circ \varphi = Id_E$.*

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 $\psi \circ \varphi = Id_E$ 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 into F. If $\varphi \circ \alpha = \varphi \circ \beta$, then $\alpha = \beta$.*

Proof. By Proposition 1.3, there exists a mapping $\psi : \mathrm{Im}(\varphi) \to E$ such that $\psi \circ \varphi = Id_F$. 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 $\varphi \circ \varphi^{-1} = Id_F$ and $\varphi^{-1} \circ \varphi = Id_E$.*

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 $\varphi^{-1} \circ \varphi = Id_E$ requires that $x = \varphi^{-1}(\varphi(x)) = \varphi^{-1}(y)$. This insures the unicity 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 $\varphi \circ \varphi^{-1} = Id_F$ and $\varphi^{-1} \circ \varphi = Id_E$. □

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

It is clear that the composition of two injective (resp. surjective) mappings is injective (resp. 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 $\varphi \circ \psi = Id_F$. 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 it is the inverse of a partial function.*

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

Suppose now that τ 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 every $X, Y \subseteq E$, $X \cap Y = \emptyset$ 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 $x = y$ 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) *$\tau^{-1} \circ \tau = Id_{\mathrm{Dom}(\tau)}$,*
(3) *$\tau^{-1} \circ \tau \subseteq Id_E$.*

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 $\tau^{-1} \circ \tau \subseteq Id_{\mathrm{Dom}(\tau)}$. But if $x \in Id_{\mathrm{Dom}(\tau)}$, there exists by definition $y \in \tau(x)$ and thus $x \in \tau^{-1} \circ \tau(x)$. Thus $\tau^{-1} \circ \tau = Id_{\mathrm{Dom}(\tau)}$.
(2) implies (3) is trivial.
(3) implies (2). Suppose that $\tau^{-1} \circ \tau \subseteq Id_E$ 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 $\tau^{-1} \circ \tau \subseteq Id_E$, 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) *$\tau \circ \tau^{-1} = Id_{\mathrm{Im}(\tau)}$,*
(3) *$\tau \circ \tau^{-1} \subseteq Id_F$.*

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 $\tau \circ \tau^{-1} = Id_F$.*

Proof. Suppose that τ is a surjective partial function and let $y \in F$. Then $y = \tau(x)$ for some $x \in E$ and thus $y \in \tau(\tau^{-1}(y))$. Furthermore, if $y' \in \tau(\tau^{-1}(y))$, then $y' = \tau(x')$ for some $x' \in \tau^{-1}(y)$. It follows that $\tau(x') = y$ and thus $y = y'$. Thus $\tau \circ \tau^{-1} = Id_F$.

Conversely, if $\tau \circ \tau^{-1} = Id_F$, then τ is a partial function by Corollary 1.11. Let $y \in F$. Then $y \in \tau \circ \tau^{-1}(y)$ and thus $y = \tau(x)$ for some $x \in \tau^{-1}(y)$. Therefore τ is surjective. □

Corollary 1.12 is often used under 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, $\beta \circ \beta^{-1} = Id_F$. Thus $\gamma \circ \beta^{-1} = \alpha$. □

It is clear that the composition of two injective (resp. surjective) relations is injective (resp. surjective). Proposition 1.6 can be also 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, 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 $z \in \alpha(x) \cap \alpha(y)$. Let $t \in \beta(z)$. 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.

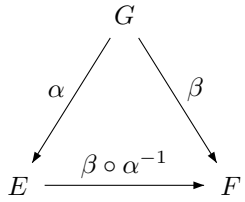
If furthermore α is a surjective partial function, then by Proposition 1.8, α^{-1} is an injective relation and by Corollary 1.12, $\alpha \circ \alpha^{-1} = Id_F$. It follows that $\gamma \circ \alpha^{-1} = \beta \circ \alpha \circ \alpha^{-1} = \beta$. Thus β is the composition of the two injective 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, $\beta^{-1} \circ \beta = Id_F$. 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 summarized 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:*

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

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 relation.

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

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

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

More precise results hold for inverse 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:*

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

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 $\varphi^{-1}(X^c) = \varphi^{-1}(F \backslash X) = \varphi^{-1}(F) \setminus \varphi^{-1}(X) = E \setminus \varphi^{-1}(X) = (\varphi^{-1}(X))^c$. □

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

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

Proof. By Corollary 1.12, $\varphi \circ \varphi^{-1} = Id_F$. It follows that $\varphi(X) \cap Y = \varphi(X) \cap \varphi\big(\varphi^{-1}(Y)\big)$. Furthermore, $\varphi\big(X \cap \varphi^{-1}(Y)\big) \subseteq \varphi(X) \cap \varphi\big(\varphi^{-1}(Y)\big)$. Finally, 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 2.1 *Any intersection of preorders (resp. equivalence relations) is a preorder (resp. an equivalence relation).*

It follows that, given a set R of relations on a set E, there is a smaller preorder (resp. equivalence relation) containing all the relations of E. This relation is called the *preorder* (resp. equivalence relation) *generated* by R.

An *order ideal* of an ordered set $(E, \leqslant)$ is a subset I of E such that, if $x \leqslant y$ and $y \in I$, then $x \in I$. The order ideal generated by an element x is the set $\downarrow x$ of all $y \in E$ such that $y \leqslant x$. The intersection (resp. union) of any family of order ideals is also an order ideal.

Chapter II

Semigroups

1 Semigroups, monoids and groups

1.1 Semigroups, monoids

Let S be a set. A *binary operation* on S is a mapping from $S \times S$ into 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 *multiplication* the binary operation. 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* or *unit* 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 a identity, and $11' = 1$, since $1'$ is a identity, whence $1 = 1'$.

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 that S is a semigroup and assume its 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 *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 (resp. monoid, group) is said to be *commutative* if its operation is commutative.

If S is a semigroup, S^1 denotes 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.3.2.

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* (resp. *left identity*) of S if, for all $s \in S$, $se = s$ (resp. $es = s$). Observe that e is an identity if and only if it is simultaneously a right and a left identity. Furthermore, a right (resp. 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* (resp. *right zero*, *left zero*) if, for all $s \in S$, $es = e = es$ (resp. $se = e$, $es = e$).

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, we denote by S^0 the semigroup obtained from S by addition of a zero: the support of S^0 is the disjoint union of S and the singleton[1] 0 and the multiplication (here denoted $*$) 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.

Cancellative elements

An element s of a semigroup S is said to be *right cancellative* (resp. *left cancellative*) if, for every $x, y \in S$, the conditions $xs = ys$ (resp. $sx = sy$) imply $x = y$. It is *cancellative* if it is simultaneously right and left cancellative.

A semigroup S is *right cancellative* (resp. *left cancellative*, *cancellative*) if all its elements are right cancellative (resp. left cancellative, cancellative).

[1] A singleton $\{s\}$ will also be denoted by s.

Inverses

We have to face a slight terminological problem with the notion of inverse. In group theory, and elsewhere in mathematics, an inverse is defined as follows. Given an element x of a monoid M, a *right inverse* (resp. *left inverse*) of x is an element x' such that $xx' = 1$ (resp. $x'x = 1$). An *inverse* of x is an element x' which is simultaneously a right and left inverse of x, so that $xx' = x'x = 1$.

Semigroup theorists have introduced a different notion with the same name. Given an element x of a semigroup S, an element x' is an *inverse* of x if $xx'x = x$ and $x'xx' = x'$.

Usually, the context will permit to clarify which definition is understood. If some ambiguity subsists, the term *group inverse* will be used for the first definition ($xx' = x'x = 1$), and *semigroup inverse* for the latter one ($xx'x = x$ 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 is given in Section V.3.

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 inverse and a left inverse.*

Proof. In a group, every element has a right inverse and a left inverse. Conversely, let G be a monoid in which every element has a right inverse and a left inverse. Let $g \in G$, let g' (resp. g'') be a right (resp. 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 groups, 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 inverse.*

Proof. Let G be a finite monoid in which every element has a left 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 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 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 inverse.*

Proof. Let G be a group. Let $g, x, y \in G$ and let g' be an 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 inverses of g,

$gg' = gg''$ and thus $g' = g''$. Thus every element has a unique inverse. □

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

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, called the *dual* of S. *Ordered monoids* are defined analogously.

1.5 Examples

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

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 $s \in S$. 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

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

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 multiplication $a_i a_j = a_j$ 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

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

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 $SL_2(\mathbb{Z})$.

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) The monoid $U_1 = \{1, 0\}$ can be ordered by setting $0 \leqslant 1$.

1.6 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 into 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 into a monoid T satisfying (1) and

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

A *morphism of ordered monoids* is a map φ from an ordered monoid $(S, \leqslant)$ into 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 S and T is a monoid morphism φ satisfying, for every $s \in S$, $\varphi(s^{-1}) = \varphi(s)^{-1}$. In fact, this condition follows from (1) and (2) since $\varphi(s^{-1})\varphi(s) = \varphi(ss^{-1}) = \varphi(1) = 1$ and similarly $\varphi(s)\varphi(s^{-1}) = \varphi(s^{-1}s) = \varphi(1) = 1$.

The semigroups (or monoids), together with the morphisms defined above, form a category. We shall encounter in Chapter X 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 $\varphi \circ \psi = Id_T$ and $\psi \circ \varphi = Id_S$.

Proposition 1.6 *In the category of semigroups (resp. monoids, groups), 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 $\varphi \circ \psi = Id_T$ and $\psi \circ \varphi = Id_S$.

Suppose now that $\varphi : S \to T$ is a bijective morphism. Then φ^{-1} is a morphism from T into 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 1.6 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.

2 Basic algebraic structures

2.1 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 2.1 *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 2.1 can be summarized as follows: *substructures are preserved by morphisms and by inverse morphisms*. A similar statement holds for monoid morphisms and for group morphims.

Given a subset R of a semigroup S, the subsemigroup of S *generated* by R is the smallest subsemigroup of S containing R. Is is denoted by $\langle R \rangle$ and consists of all products $r_1 \cdots r_n$ of elements of R. If S is a monoid, the *submonoid generated* by R is defined in a similar way, but it always contains the identity of S. Finally, if S is a group, the *subgroup generated* by R is the smallest subgroup of S containing R. It consists of all products of the form $r_1 \cdots r_n$, where each r_i is either an element of R or the inverse of an element of R.

A semigroup (resp. monoid, group) is called *monogenic* if it is generated by a single element.

2.2 Quotients, divisions and products

Let S and T be two semigroups (resp. monoids). Then T is a *quotient* of S if there exists a surjective morphism from S onto T. Finally, a semigroup S *divides* a semigroup T (notation $S \preccurlyeq T$) and if S is quotient of a subsemigroup of T.

Proposition 2.2 *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 2.3 *Two finite semigroups that divide each other are isomorphic.*

Proof. We keep the notation of the proof of Proposition 2.2, 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. □

Given a family $(S_i)_{i\in I}$ of semigroups (resp. monoids), the *product* $\Pi_{i\in I} S_i$ is the semigroup (monoid) 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 monoid 1 is the identity for the product of semigroups (resp. monoids). Following an 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 $(M_i)_{i\in I}$.

2.3 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 to include the cases $x = 1$ and $y = 1$, which are necessary to recover the conditions $SI \subseteq S$ and $SI \subseteq I$. 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 (resp. right ideals, left ideals) of S is again an ideal (resp. right ideal, left ideal).

Let R be a subset of a semigroup S. The ideal (resp. right ideal, left ideal) *generated* by R is the set S^1RS^1 (resp. RS^1, S^1R). It is the smallest ideal (resp. right ideal, left ideal) containing R. An ideal (resp. right ideal, left ideal) is called *principal* if it is generated by a single element. Note that the ideal (resp. right ideal, left ideal) generated by an idempotent e is equal to SeS (resp. eS, Se). Indeed, the equality $S^1eS^1 = SeS$ follows from the fact that $e = eee$.

Ideals are stable under surjective morphisms and inverse of morphisms.

Proposition 2.4 *Let $\varphi : S \to T$ be a semigroup morphism. If J is an ideal of T, then $\varphi^{-1}(J)$ is a subsemigroup 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

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

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

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

is the *product* of the ideals $I_1, \ldots, I_n$.

Proposition 2.5 *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 2.6 *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 2.5, 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, $I = I_1 = I_2$. □

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 2.1 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.

2.4 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 2.7 *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 2.8

(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 $s \in S \setminus 0$.*

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 2.7 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 $s \in S \setminus 0$, we have $S = SsS \subseteq S^2$ and therefore $S^2 \neq 0$. Furthermore, 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 simple semigroups will be detailed in Section V.4.

2.5 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 a structure of 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 $(S \setminus I) \cup 0$ and the multiplication (here denoted $*$) is defined by

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

(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 $x, y \in S^1$,

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

This congruence 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 $S \times S$. Further, it is closed under intersection. It follows that the intersection of all congruences containing R is a congruence, called the *congruence generated by* R.

The proposition below gives a more constructive definition.

Proposition 2.9 *The congruence generated by a relation R on a semigroup S is the reflexive-transitive closure of the relation $\{(xry, xsy) \mid (r, s) \in R\}$.*

Proof. TO DO.

(d) *Nuclear congruence*

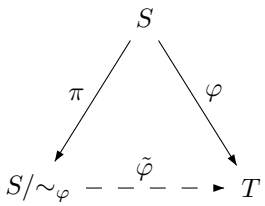
For each semigroup morphism $\varphi : S \to T$, the equivalence $\sim_\varphi$ defined on S by

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

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

Proposition 2.10 (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 $\bar{\varphi} : S/\sim_\varphi \to T$ such that $\varphi = \bar{\varphi} \circ \pi$. Moreover, $\bar{\varphi}$ is an isomorphism from $S/\sim_\varphi$ onto $\varphi(S)$.*

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



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

$$\tilde{\varphi}(s) = \varphi(x) \tag{2.1}$$

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 (2.1). Moreover, if $\pi(x_1) = s_1$ and $\pi_2(x_2) = s_2$, 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 $\bar{\varphi}$ 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 2.11 (Second isomorphism theorem) *Let $\sim_1$ and $\sim_2$ be two congruences on a semigroup S and π_1 (resp. π_2) the canonical morphism from S onto $S/\!\sim_1$ (resp. $S/\!\sim_2$). 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 2.12 *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 2.10, $S/\!\sim$ is isomorphic to $\pi(S)$. □

3 Transformation semigroups

3.1 Definitions

A *transformation* on a set P is a map from P into itself. A *permutation* is a bijective transformation.

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 $(p, s) \mapsto p\cdot s$, 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 faithfull 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 3.1 Each semigroup S defines a transformation semigroup (S^1, S), given by the faithful action $q\cdot s = qs$.

Example 3.2 With each finite set R, one can associate 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 *fixpoint-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 fixpoint-free transformation semigroup.

3.2 Full transformation semigroup and symmetric group

The *full transformation semigroup* on a set P is the semigroup $\mathfrak{T}(P)$ of all transformations on P. If $P = \{1,\ldots,n\}$, the notation $\mathfrak{T}_n$ 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 $\mathfrak{T}(P)$ is reverse composition.

The *symmetric group* on a set P is the group $\mathfrak{S}(P)$ of all permutations on P. If $P = \{1,\ldots,n\}$, the notation $\mathfrak{S}_n$ is also used.

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

Proposition 3.1 *Every semigroup S is isomorphic to a subsemigroup of $\mathfrak{T}(S^1)$. In particular, every finite semigroup is isomorphic to a subsemigroup of $\mathfrak{T}_n$ 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 into $\mathfrak{T}(S^1)$ and thus S is isomorphic to a subsemigroup of $\mathfrak{T}(S^1)$. □

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

Theorem 3.2 (Cayley theorem) *Every group G is isomorphic to a subgroup of $\mathfrak{S}(G)$. In particular, every finite group is isomorphic to a subgroup of $\mathfrak{S}_n$ for some n.*

3.3 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 3.3 *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 $s_1 = s_2$. Therefore, there is a well-defined map $\pi : \hat{s} \to s$ 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). □

4 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 by mere juxtaposition

$$a_0a_1\cdots a_n$$

the sequence $(a_0, a_1, \ldots, a_n)$. The set of words is endowed with the operation of *concatenation product* also called *product*, which associates with two words $x = a_0a_1\cdots a_p$ and $y = b_0b_1\cdots b_q$ the word $xy = a_0a_1\cdots a_pb_0b_1\cdots b_q$. This operation is associative. It has an identity, the *empty word*, denoted by 1 or ε and which is the empty sequence.

If u is a word and a a letter, we denote by $|u|_a$ the number of occurrences of a in u. Thus, if $A = \{a, b\}$ and and $u = abaab$, we have $|u|_a = 3$ and $|u|_b = 2$. The sum

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

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

We denote by A^* the set of words on A and by A^+ the set of nonempty words. The set A^* (resp. A^+), equipped with the concatenation product is thus a monoid with identity 1 (resp. a semigroup). The set A^* is called the *free monoid* on A and A^+ the *free semigroup* on A. Free structures are defined in category theory by a so-called universal property. The next proposition, which shows that A^+ (resp. A^*) satisfies this universal property, justifies our terminology.

Proposition 4.1 *If φ is a function from A into a semigroup (resp. monoid) S, there exists a unique semigroup (resp. monoid) morphism $\bar{\varphi} : A^+ \to S$ (resp. $A^* \to S$) 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 $a_0 a_1 \cdots a_n$,

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

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 these two equalities, which shows it is unique.

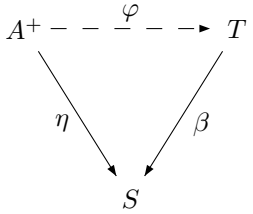
By construction, the set $\varphi(A)$ generates $\bar{\varphi}(A)$. Consequently, the morphism $\bar{\varphi}$ is surjective if and only if the set $\varphi(A)$ generates S. □

These results have several frequently used corollaries.

Corollary 4.2 *Let S be a semigroup and let A be a subset of S generating S. The identity map from A into S induces a morphism of semigroups from A^+ onto S.*

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

Corollary 4.3 *Let $\eta : A^+ \to S$ be a morphism and $\beta : T \to S$ be 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 4.1, can be extended to a morphism $\varphi : A^+ \to T$ such that $\eta = \beta \circ \varphi$. □

5 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 strictly 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 5.1.

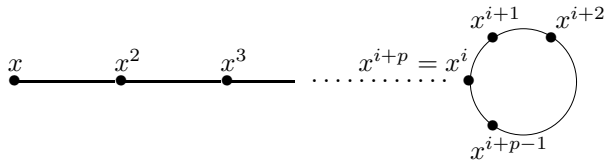


Figure 5.1. The semigroup generated by x.

The next result is a key property of finite semigroups.

Proposition 5.1 *In a finite semigroup, each element 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, we have

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

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 5.1 has two important consequences.

Corollary 5.2 *Every nonempty finite semigroup contains at least one idempotent.*

Proposition 5.3 *For each finite semigroup S, there exists an integer ω such that, for each $s \in S$, s^ω is idempotent.*

Proof. By Proposition 5.1, 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 5.3 is called the *exponent* of S.

Here is another elementary property connected with idempotents.

Proposition 5.4 *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_1s_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 5.5 *Let S be a finite semigroup and let $E(S)$ be the set of idempotents of S. Then for every $n \geqslant |S|$, $S^n = SE(S)S$.*

Let us state a very useful result on finite groups.

Proposition 5.6 *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 5.1, s has an idempotent power, which is necessarily the identity of G. 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. □

We already stated a combinatorial property of finite semigroups: Proposition 5.4 states that every sufficiently long sequence of elements of a finite semigroup contains a subsequence of consecutive elements whose product is idempotent. The proof of this result was elementary and relied mainly on the pigeon-hole principle. 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.

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

Theorem 5.7 (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 having at least N elements and for every colouring in m colours 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 generalizes Proposition 5.4.

Theorem 5.8 *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 factorization $w = xu_1 \cdots u_k y$ with $x, y \in A^*$, $u_1, \ldots, u_k \in A^+$ 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 into $|S|$ colours of pairs of elements of $\{1, \ldots, N\}$ 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

$$\varphi(a_{i_0}) \cdots \varphi(a_{i_1-1}) = \varphi(a_{i_1}) \cdots \varphi(a_{i_2-1}) = \varphi(a_{i_0}) \cdots \varphi(a_{i_2-1})$$

whereby $ee = e$. Thus e is idempotent and we obtain the required factorization 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 5.8, but their order is important. In particular, the integer N does not depend on the size of A, which can even be infinite.

Theorem 5.9 *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 factorization $w = xu_1 \cdots u_k y$ with $x, y \in A^*$, $|u_1| = \ldots = |u_k|$ and $\varphi(u_1) = \ldots = \varphi(u_k) = e$.*

6 Exercises.

Exercice 1 Let T be a semigroup and let R and S be subsemigroups of T. Show that $R \cup S$ is a subsemigroup of T if and only if $ST \cup TS$ is a subset of $S \cup T$.

Show that this condition is satisfied if S and T are both left ideals or both right ideals, or if either S or T is an ideal.

Exercice 2 Let (P, S) be a transformation semigroup. Show that (P, S) divides $(P, 1_P) \times (S^1, S)$.

Chapter III

Languages and automata

This chapter is a brief overview of the theory of finite automata and formal languages. For a complete introduction to this theory, the reader is referred to specialised books.

There are different manners to describe a set of words, or *languages*. 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, given in Section 2 is of this type. In the descriptive approach, the words of a language are characterized by a property: the language of words of even length, the set of binary representation of prime numbers are examples of this approach. The machine approach is a special case of the descriptive approach: a machine reads a word as input and decides is the word is accepted or not. The set of words accepted by the machine defines a language.

1 Langages

Let A be a finite alphabet. The 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:

$$\begin{aligned} L \cup L' &= \{u \in A^* \mid u \in L \text{ ou } u \in L'\} \\ L \cap L' &= \{u \in A^* \mid u \in L \text{ et } u \in L'\} \\ L^c &= A^* \setminus L = \{u \in A^* \mid u \notin L\} \\ L \setminus L' &= L \cap L'^c = \{u \in A^* \mid u \in L \text{ et } u \notin L'\}. \end{aligned}$$

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

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

The concatenation product is an associative operation on the set of languages,

which admits the language[1] $\{1\}$ as an identity, since one has for each language L

$$\{1\}L = L\{1\} = L$$

Thus the languages over A^* form a monoid for the product. Note that this monoid is not commutative if the alphabet contains at least two letters. The product is distributive over union, which means that, for all languages L, L_1 and L_2, one has

$$L(L_1 \cup L_2) = LL_1 \cup LL_2$$

However, it is not distributive over intersection. For instance

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

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 union of all the powers of L:

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

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

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

Example 1.1 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^* (resp. A^+) is compatible with the definition of the operations L^* and L^+. Also note the following formula

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

To avoid too much parentheses, it is convenient to define precedence orders for operators on languages, summarizes in the following table 1.1.

Operateur	**Priorit**
L^*, L^c	1
L_1L_2	2
$L_1 \cup L_2$, $L_1 \cap L_2$	3

Table 1.1. Operation precedence table.

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

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

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 $\emptyset$ and $\{a\}$ for each $a \in A$,

(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 $L \cup L'$, 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 the conditions (a) and (b). Further, 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 $\emptyset^* = \{1\}$. 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,

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

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

Proposition 2.1 *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 $K \cup K'$, 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 (resp. odd) length is rational. Indeed, this language can be written as $(A^2)^*$ (resp. $(A^2)^*A$).

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 $(e \cup e')$, (ee') and e^* are rational expressions.

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

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

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

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

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

The difficulty raised by this example is deeper than it seems. Even if a rational language can be represented by infinitely many different raional expressions, one could expect to have a unique reduced expression, up to a set of simple identities like $(L^*)^* = L^*$, $L \cup K = K \cup L$ or $L(K \cup K') = LK \cup LK'$. 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 [5, 14].

3 Finite automata

A finite *automaton (fini)* 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*, 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 the diagram 3.1.

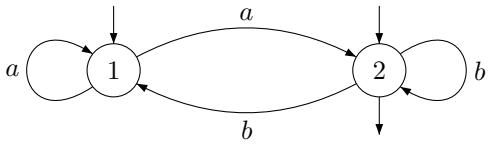


Figure 3.1. An automaton.

Let (p, a, q) be a transition: p is the *origin*, q the *end* and a the *label* of the transition. 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 by

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

Chapter IV

Recognizable and rational sets

The notions of rational and recognizable sets are usually defined for free monoids, under the common name of regular sets.

1 Rational subsets of a monoid

Let M be a monoid. The product of two subsets X and Y of M is the subset

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

Union and product grant the set $\mathcal{P}(M)$ of subsets of M with a structure of semiring. For this reason, we shall frequently denote the union additively, and the empty set by 0. The identity for the multiplication is the singleton $\{1\}$, where 1 is the identity of M. Since the map $m \to \{m\}$ is an embedding, one can identify M with a submonoid of the multiplicative monoid $\mathcal{P}(M)$. This leads to the convenient abuse of notation consisting in denoting simply by m the singleton $\{m\}$.

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:

$$X^* = \sum_{n \geqslant 0} X^n = 1 + X + X^2 + X^3 + \cdots$$
$$X^+ = \sum_{n > 0} X^n = X + X^2 + X^3 + \cdots$$

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 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

$$\{v_0 + n_1 v_1 + \dots n_r v_r \mid n_1, \dots, n_r \in \mathbb{N}\}$$

where v_0, v_1, ..., v_r are vectors of $\mathbb{N}^k$.

The rational subsets are by construction closed under union, product and star. They are also stable under morphism.

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 M, 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. Further, 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 it is well-known that this language is not rational.

It follows also 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.

Rational subsets are closed under direct products, in the following sense:

Theorem 1.2 *Let R_1 (resp. R_2) be a rational subset of a monoid M_1 (resp. 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 1.3 *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 Recognizable subsets of a monoid

Recognizable languages are usually defined in terms of automata. This is the best definition from an algorithmic point of view, but it is an asymmetric notion. It turns out that to handle the fine structure of recognizable languages, it is more appropriate to use a more abstract definition, using monoids in place of automata, due to Rabin and Scott [30]. Although these definitions will be mainly used in the context of free monoids, it is as simple to give them in a more general setting.

2.1 Recognition by morphisms

Let $\varphi : M \to N$ be a monoid morphism. A subset L of M is *recognized* by φ if there exists a subset P of N such that

$$L = \varphi^{-1}(P)$$

Let us start by an elementary, but useful observation:

Proposition 2.1 *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 recognized by φ,*
(2) *L is saturated by $\sim_\varphi$,*
(3) $\varphi^{-1}(\varphi(L)) = L$.

Proof. (1) implies (2). If L is recognized 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 recognized by φ. □

Note that one can always assume that φ is surjective. Indeed φ induces a morphism from M onto $\varphi(N)$ and $Q = \varphi^{-1}(P \cap \varphi(M))$. Furthermore, if φ is surjective, Proposition I.1.19 shows that the conditions $L = \varphi^{-1}(P)$ and $P = \varphi(L)$ are equivalent.

By extension, one also says that a monoid N recognizes a subset L of a monoid M if there exists a monoid morphism $\varphi : M \to N$ that recognizes 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 form $(\mathbb{N}, +)$ onto T defined by $\varphi(0) = 0$, $\varphi(1) = 1$ and $\varphi(n) = 2$ for all $n \geqslant 2$. The subsets of $\mathbb{N}$ recognized 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 $aba = b$, $bab = b$, $aa = bb = 0$. 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 $\varphi^{-1}(0) = A^*aaA^* \cup A^*bbA^*$.

2.2 Connection with automata

The case of the free monoid is of course the most important. In this case, our definition is equivalent with the standard definition using automata.

Recall that a (nondeterministic) *automaton* is a quintuple $\mathcal{A} = (Q, A, E, I, F)$, where A denotes a finite alphabet, Q is the set of *states*, E is the set of *transitions* (a subset of $Q \times A \times Q$), and I and F are the set of *initial* and *final states*, respectively. An automaton $\mathcal{A} = (Q, A, E, I, F)$ is *deterministic* if I is a singleton and if the conditions $(p, a, q), (p, a, q') \in E$ imply $q = q'$. An automaton is finite if its set of states is finite.

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

$$e_0 = (q_0, a_0, q_1), \quad e_1 = (q_1, a_1, q_2), \quad \ldots \quad , \quad e_{n-1} = (q_{n-1}, a_{n-1}, q_n)$$

also denoted

$$q_0 \xrightarrow{a_0} q_1 \xrightarrow{a_1} q_2 \cdots q_{n-1} \xrightarrow{a_{n-1}} q_n$$

The state q_0 is the *origin* of the path, the state q_n is its *end*, and the word $x = a_0a_1 \cdots a_{n-1}$ is its *label*. It is convenient to have also, for each state q, an empty path of label 1 from q to q. A path in $\mathcal{A}$ is *successful* if its origin is in I and its end is in F. The language of A^* *recognized* by $\mathcal{A}$ is the set of the labels of all successful paths of $\mathcal{A}$.

Automata are conveniently represented by labeled graphs, as shown in Figure 2.1. Incoming arrows indicate initial states and outgoing arrows indicate final states.

Example 2.3 Let $\mathcal{A} = (Q, A, E, I, F)$ be the automaton represented below, with $Q = \{1, 2\}$, $A = \{a, b\}$, $I = \{1\}$, $F = \{2\}$ and

$$E = \{(1, a, 1), (1, a, 2), (2, a, 2), (2, b, 2), (2, b, 1)\}.$$

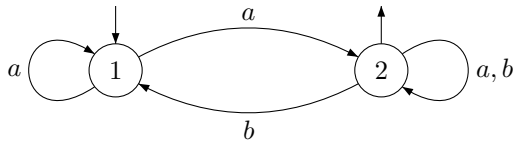


Figure 2.1. A nondeterministic automaton.

The path $(1, a, 1)(1, a, 2)(2, b, 2)$ is a successful path of label aab. The path $(1, a, 1)(1, a, 2)(2, b, 1)$ has the same label but is unsuccessful since its end is 1. The set of words accepted by $\mathcal{A}$ is aA^*, the set of all words whose first letter is a.

The equivalence between automata and monoids is based on the following observation. Let $\mathcal{A} = (Q, A, E, I, F)$ be a 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. It is not difficult to see that μ is a monoid morphism from A^* into the monoid of relations on Q. The monoid $\mu(A^*)$ is called the *transition monoid* of $\mathcal{A}$, 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

$$\mu(u)_{p,q} = \begin{cases} 1 & \text{if there exists a path from } p \text{ to } q \text{ with label } u \\ 0 & \text{otherwise} \end{cases}$$

Note that a word u is recognized by $\mathcal{A}$ if and only if $(p, q) \in \mu(u)$ for some initial state p and some final state q. This leads to the next proposition.

Proposition 2.2 *If a finite automaton recognizes a language L, then its transition monoid recognizes L.*

Example 2.4 If $\mathcal{A} = (Q, A, E, I, F)$ is the automaton of example 2.3, 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

$$\mu(A^*) = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \right\}.$$

The previous computation can be simplified if $\mathcal{A}$ is deterministic. Indeed, in this case, the transition monoid of $\mathcal{A}$ is naturally embedded into the monoid of partial transformations on Q.

Example 2.5 Let $A = \{a, b\}$ and let $\mathcal{A}$ be the (incomplete) deterministic automaton represented below.

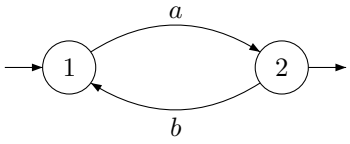


Figure 2.2. A deterministic automaton.

It is easy to see that $\mathcal{A}$ recognizes the language $A^* \setminus (ab)^*$. 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 0. The other relations defining S are $aba = a$, $bab = b$ and $bb = 0$.

	a	b	aa	ab	ba
1	2	–	–	1	–
2	–	1	–	–	2

One recognizes the monoid B_2^1.

Conversely, given a monoid morphism $\varphi : A^* \to M$ and a subset P of M, one can build a deterministic automaton recognizing $L = \varphi^{-1}(P)$ as follows. Take the right representation of A on M defined by $s \cdot a = s\varphi(a)$. This defines an automaton $\mathcal{A} = (M, A, E, \{1\}, P)$, where $E = \{(s, a, s \cdot a) \mid s \in M, a \in A\}$ that recognizes L.

Example 2.6 Let $\varphi : \{a, b\}^* \to B_2^1 = \{1, a, b, ab, ba, 0\}$ be the morphism defined by $\varphi(a) = a$ and $\varphi(b) = b$. By applying the algorithm described above, one gets the automaton pictured in Figure 2.3, which also recognizes $(ab)^*$.

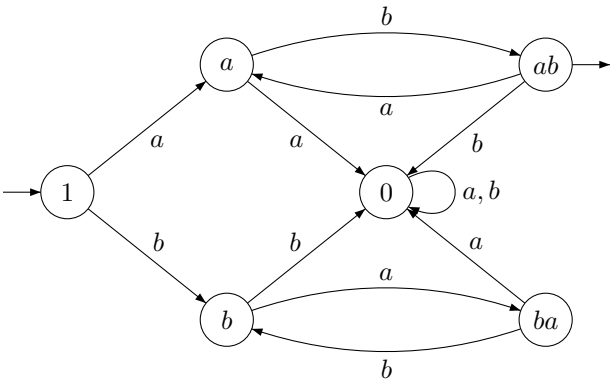


Figure 2.3. The automaton associated with φ.

2.3 Operations on sets

Simple operations on sets have a natural algebraic counterpart. We now study in this order complement, intersection, union, inverse morphisms and left and right quotients.

Proposition 2.3 *Let L be a subset of the monoid M. If L is recognized $\varphi : M \to N$, then $M \setminus L$ is also recognized by φ.*

Proof. If $L = \varphi^{-1}(P)$ then, by Proposition I.1.18, $M \setminus L = \varphi^{-1}(N \setminus P)$. □

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 $\varphi(x_1, \ldots, x_n) = (\varphi_1(x_1), \ldots, \varphi_n(x_n))$.

Proposition 2.4 *Let $L_1, \ldots, L_n$ be subsets of M. If each L_i is recognized by φ_i, then the sets $\cap_{1 \leqslant i \leqslant n} L_i$ and $\cup_{1 \leqslant i \leqslant n} L_i$ are recognized by φ.*

Proof. Suppose that $L_i = \varphi^{-1}(P_i)$ 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 2.5 *Let $\eta : R \to M$ and $\varphi : M \to N$ be two surjective morphisms of monoids. If φ recognizes a subset L of M, then $\varphi \circ \eta$ recognizes $\varphi^{-1}(L)$.*

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 $\varphi \circ \eta$ recognizes $\varphi^{-1}(L)$. □

Recall that, for each subset X of S and for each element s of M, the *left* (resp. *right*) *quotient* $s^{-1}X$ (resp. Xs^{-1}) of X by s is defined as follows:

$$s^{-1}X = \{t \in S \mid st \in X\} \quad \text{and} \quad Xs^{-1} = \{t \in S \mid ts \in X\}$$

More generally, for any subset K of M, the left (resp. right) quotient $K^{-1}X$ (resp. XK^{-1}) of X by K is

$$K^{-1}X = \bigcup_{s \in K} s^{-1}X = \{t \in S \mid \text{ there exists } s \in K \text{ such that } st \in X\}$$
$$XK^{-1} = \bigcup_{s \in K} Xs^{-1} = \{t \in S \mid \text{ there exists } s \in K \text{ such that } ts \in X\}$$

Proposition 2.6 *Let $\varphi : M \to N$ be a surjective morphism of monoids. If φ recognizes a subset L of M, it also recognizes $K^{-1}L$ 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 recognizes $K^{-1}L$. A similar proof works for LK^{-1}. □

2.4 Recognizable sets

A subset of a monoid is *recognizable* if it is recognized by a finite monoid. We denote by $\mathrm{Rec}(M)$ the set of recognizable subsets of M.

Propositions 2.4, 2.5 and 2.6 give immediately

Corollary 2.7 *For any monoid M, $\mathrm{Rec}(M)$ is closed under Boolean operations and left and right quotients. Further, if $\eta : M \to N$ is a surjective morphism, $L \in \mathrm{Rec}(M)$ implies $\varphi^{-1}(L) \in \mathrm{Rec}(N)$.*

If $M = A^*$, where A is a finite alphabet, the results of Section 2.2 show that a language is recognizable if and only if it is recognized by a finite automaton.

2.5 Recognition by ordered monoids

We shall now slightly modify the standard definition by introducing ordered monoids. This order occurs quite naturally and permits to distinguish between a language and its complement.

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, 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 $M/\!\preccurlyeq$.

Let $\varphi : M \to N$ be a surjective morphism of ordered monoids. A subset Q of M is *recognized* by φ if there exists an order ideal P of N such that

$$Q = \varphi^{-1}(P)$$

This condition implies that Q is an order ideal of M and that $\varphi(Q) = \varphi(\varphi^{-1}(P)) = P$. By extension, a subset Q of M is said to be *recognized* by an ordered monoid N if there exists a surjective morphism of ordered monoids from M onto N that recognizes 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 *recognized* by $\preccurlyeq$ if, for every $q \in Q$, $p \preccurlyeq q$ implies $p \in Q$. It is easy to see that a surjective morphism of ordered monoids φ recognizes Q if and only if the nuclear congruence $\preccurlyeq_\varphi$ recognizes Q.

2.6 Syntactic order

The syntactic congruence is one of the key notions of this chapter. Roughly speaking, it is the monoid analog of the notion of minimal automaton. First note that, if M is an ordered monoid, the congruence $\leqslant$ recognizes every order ideal of M. The syntactic congruence of an order ideal Q of M is the coarsest congruence among the congruences on M that recognize Q.

Let M be an ordered monoid and let P be an order ideal of M. Define a relation $\preccurlyeq_P$ on M by setting

$$u \preccurlyeq_P v \text{ if and only if, for every } x, y \in M,\ xvy \in P \Rightarrow xuy \in P$$

One can show that the relation $\preccurlyeq_P$ is a congruence of ordered monoids on M that recognizes P. This congruence is called the *syntactic congruence* of P in M. The equivalence relation associated with $\preccurlyeq_P$ is denoted $\sim_P$ and called the *syntactic equivalence* of P in M. Thus $u \sim_P v$ if and only if, for every $x, y \in M$,

$$xuy \in P \Longleftrightarrow xvy \in P$$

The ordered monoid $M(P) = M/\!\preccurlyeq_P$ 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 characterized by the following property.

Proposition 2.8 *The syntactic congruence of P is the coarsest congruence that recognizes P. Furthermore, a congruence $\preccurlyeq$ recognizes P if and only if $\preccurlyeq_P$ is coarser than $\preccurlyeq$.*

It is sometimes convenient to state this result in terms of morphisms:

Corollary 2.9 *Let $\varphi : M \to N$ be a surjective morphism of ordered monoids and let P be an order ideal of M. The following properties hold:*

(1) *The morphism φ recognizes P if and only if η_P factorizes through it.*

(2) *Let $\pi : N \to M$ be a surjective morphism of ordered monoids. If $\pi \circ \varphi$ recognizes P, then φ recognizes P.*

2.7 How to compute the syntactic monoid?

The easiest way to compute the syntactic ordered monoid of a recognizable language L is to first compute its minimal (deterministic) automaton $\mathcal{A} = (Q, A, \cdot, \{q_0\}, F)$. Then the syntactic monoid of L is equal to the transition monoid M of $\mathcal{A}$ and the order on S is given by $s \leqslant t$ if and only if,

$$\text{for every } x \in M, \text{ for every } q \in Q,\ q \cdot tx \in F \Rightarrow q \cdot sx \in F$$

Example 2.7 Let $\mathcal{A}$ be the deterministic automaton of example 2.5. It is the minimal automaton of $L = A^+ \setminus (ab)^*$. The transition monoid was calculated in the previous section. The syntactic order is given by $0 \leqslant s$ for every $s \in S$. Indeed, $q \cdot 0 = 3 \in F$ and thus, the formal implication

$$q \cdot sx \in F \Rightarrow q \cdot 0x \in F$$

holds for any $q \in Q$, $s \in S$ and $x \in S^1$. One can verify that there is no other relations among the elements of S. For instance, a and ab are incomparable since $1 \cdot aa = 3$ but $1 \cdot aba = 2 \notin F$ and $1 \cdot abb = 3$ but $1 \cdot ab = 1 \notin F$.

Chapter V

Green's relations and local theory

In this chapter, all semigroups are finite.

1 Green's relations

These fundamental equivalence relations were introduced and studied by Green in 1951. They are now basic in the theory of semigroups.

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

$$\begin{aligned}
s \leqslant_{\mathcal{R}} t &\iff sS^1 \subseteq tS^1\\
s \leqslant_{\mathcal{L}} t &\iff S^1s \subseteq S^1t\\
s \leqslant_{\mathcal{J}} t &\iff S^1sS^1 \subseteq S^1tS^1\\
s \leqslant_{\mathcal{H}} t &\iff s \leqslant_{\mathcal{R}} t \text{ and } s \leqslant_{\mathcal{L}} t
\end{aligned}$$

Thus $s \leqslant_{\mathcal{J}} t$ (resp. $s \leqslant_{\mathcal{R}} t$, $s \leqslant_{\mathcal{L}} t$) if the ideal (resp. right ideal, left ideal) generated by s is contained in the ideal (resp. right ideal, left ideal) generated by t.

The equivalences associated with these four preorder relations are denoted

by $\mathcal{R}$, $\mathcal{L}$, $\mathcal{J}$ and $\mathcal{H}$, respectively. Therefore

$$\begin{aligned} s \,\mathcal{R}\, t &\iff sS^1 = tS^1 \\ s \,\mathcal{L}\, t &\iff S^1s = S^1t \\ s \,\mathcal{J}\, t &\iff S^1sS^1 = S^1tS^1 \\ s \,\mathcal{H}\, t &\iff s \,\mathcal{R}\, t \text{ and } s \,\mathcal{L}\, t \end{aligned}$$

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 (resp. $\mathcal{L}$-class, $\mathcal{J}$-class, $\mathcal{H}$-class) is denoted by $R(s)$ (resp. $L(s)$, $J(s)$, $H(s)$).

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 $s \not\mathcal{K}\, t$. The next propositions summarize some useful properties of Green's relations.

Proposition 1.1 *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 1.2 *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 $s \,\mathcal{R}\, sx \,\mathcal{R}\, sxy$. □

There is of course a dual statement for the relation $\leqslant_{\mathcal{L}}$. The first part of Proposition 1.2 can be extended to the preorder $\leqslant_{\mathcal{H}}$.

Proposition 1.3 *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. The equivalence of (1) and (2) follows from Proposition 1.2 and its dual version for $\leqslant_{\mathcal{L}}$. □

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 1.4 *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) $efe = e$.

Proof. The equivalence of (1) and (2) follows from Proposition 1.3 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 1.5 *In each semigroup S, the relations $\leqslant_{\mathcal{R}}$ and $\leqslant_{\mathcal{L}}$ (resp. $\mathcal{R}$ 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 1.5.

Proposition 1.6 *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 least preorder containing both $\leqslant_{\mathcal{R}}$ and $\leqslant_{\mathcal{L}}$.*

Proof. If $s \leqslant_{\mathcal{L}} \circ \leqslant_{\mathcal{R}} t$ then for some $r \in S$, $s \leqslant_{\mathcal{R}} r \leqslant_{\mathcal{L}} t$, whence $s \leqslant_{\mathcal{J}} r \leqslant_{\mathcal{J}} t$ and $s \leqslant_{\mathcal{J}} t$. Conversely, if $s \leqslant_{\mathcal{J}} t$, then $s = utv$ for some $u, v \in S^1$, whence $s \leqslant_{\mathcal{R}} ut \leqslant_{\mathcal{L}} t$. Thus $\leqslant_{\mathcal{J}}$ is equal to $\leqslant_{\mathcal{L}} \circ \leqslant_{\mathcal{R}}$.

Let $\leqslant_{\mathcal{D}}$ be the least preorder containing both $\leqslant_{\mathcal{R}}$ and $\leqslant_{\mathcal{L}}$. Since $\leqslant_{\mathcal{J}}$ is a preorder containing $\leqslant_{\mathcal{R}}$ and $\leqslant_{\mathcal{L}}$, it contains $\leqslant_{\mathcal{D}}$. Furthermore

$$\leqslant_{\mathcal{J}} = \leqslant_{\mathcal{L}} \circ \leqslant_{\mathcal{R}} \subseteq \leqslant_{\mathcal{D}} \circ \leqslant_{\mathcal{D}} = \leqslant_{\mathcal{D}}$$

and thus $\leqslant_{\mathcal{J}} = \leqslant_{\mathcal{D}}$. □

We now introduce the fifth Green's relation. The relation $\mathcal{D}$ is the least equivalence relation containing both $\mathcal{R}$ and $\mathcal{L}$. Proposition 1.5 immediately leads to an easier definition.

Proposition 1.7 *The relation $\mathcal{D}$ is equal to $\mathcal{L} \circ \mathcal{R}$ and to $\mathcal{R} \circ \mathcal{L}$.*

Proof. Let $\mathcal{C} = \mathcal{L} \circ \mathcal{R}$. We claim that $\mathcal{R}$ is an equivalence relation. First, it is clearly reflexive. It is also symmetric since $\mathcal{L}$ and $\mathcal{R}$ commute. Finally, it is transitive, since

$$(\mathcal{L} \circ \mathcal{R}) \circ (\mathcal{L} \circ \mathcal{R}) = \mathcal{L} \circ (\mathcal{R} \circ \mathcal{L}) \circ \mathcal{R} = \mathcal{L} \circ (\mathcal{L} \circ \mathcal{R}) \circ \mathcal{R} = (\mathcal{L} \circ \mathcal{L}) \circ (\mathcal{R} \circ \mathcal{R}) = \mathcal{L} \circ \mathcal{R}$$

Since $\mathcal{C}$ is an equivalence relation containing both $\mathcal{R}$ and $\mathcal{L}$, it contains $\mathcal{D}$, which is the least equivalence relation having this property. On the other hand,

$$\mathcal{C} = \mathcal{L} \circ \mathcal{R} \subseteq \mathcal{D} \circ \mathcal{D} = \mathcal{D}$$

and thus $\mathcal{C} = \mathcal{D}$. □

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 tempting to guess, in view of Proposition 1.6, that $\mathcal{D} = \mathcal{J}$. This equality does not hold in general for infinite semigroups (see Example 1.1 below) but it holds for finite semigroups.

Theorem 1.8 *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.5.3, 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.5.3, 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 1.1 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 1.5 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 1.1. 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.

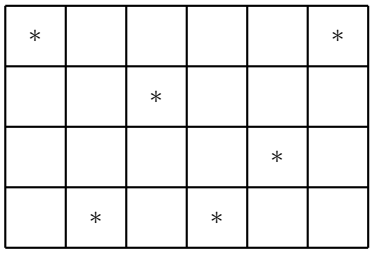


Figure 1.1. A $\mathcal{D}$-class.

The next proposition describes the structure of a $\mathcal{D}$-class.

Proposition 1.9 (Green's lemma) *Let D be a $\mathcal{D}$-class of a semigroup S, and let s and t be two $\mathcal{R}$-equivalent elements of S. If $s = tp$ and $t = sq$ for some $p, q \in S^1$, the maps $x \to xp$ and $x \to xq$ define inverse bijections between $L(s)$ and $L(t)$, and these bijections preserve the $\mathcal{H}$-classes.*

Proof. Let $n \in L(s)$ (see Figure 1.2). 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 1.1 shows that the maps $x \to xp$ and $x \to xq$ preserve the $\mathcal{H}$-classes. □

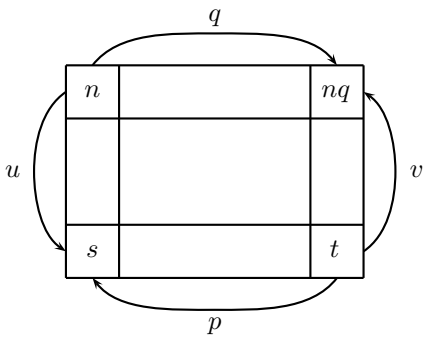


Figure 1.2. An illustration of Green's lemma.

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 1.10 (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:*

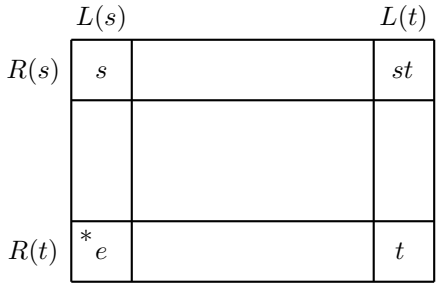
(1) $st \in R(s) \cap L(t)$,

(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 these conditions are satisfied, then $st \in D$.

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 $et = t$. Therefore $e = tv$ for some $v \in S^1$ and $ee = etv = tv = e$. Thus e is idempotent.



(2) implies (3). Let e be an idempotent of $R(t) \cap L(s)$. 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 condition (1) is satisfied, then $st \in D$, since $R(s)$ is contained in D. □

Here is a useful consequence of the Location theorem.

Proposition 1.11 *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 1.2 and $ru = e$ for some $u \in S^1$. It follows that uer is idempotent, since $ueruer = ue(ru)er = ueer = uer$. Furthermore $r(uer) = eer = er = r$. Consequently $r \mathcal{L} uer$, $L(s) = L(r) = L(uer)$ and thus the $\mathcal{L}$-class of s contains an idempotent. □

Proposition 1.12 *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.*

Proof. The equivalence of (1) and (2) follows from Theorem 1.10. 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 1.2. Finally, for each $s \in H$, the property (1) shows that 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. □

The following is another remarkable consequence of Green's lemma.

Proposition 1.13 *Two maximal subgroups of a $\mathcal{D}$-class are isomorphic.*

Proof. From Proposition 1.12, 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 $tes = ts = f$.

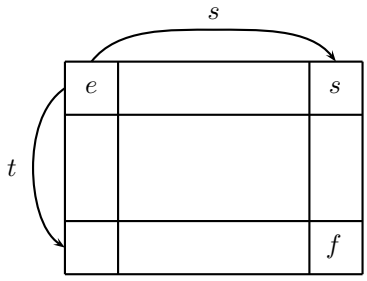


Figure 1.3. The $\mathcal{D}$-class D.

We claim that φ is a morphism. First observe that st is idempotent, since $stst = sft = st$. 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 1.2, $(st)y = y$. 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. □

2 Green's relations in finite semigroups

For a finite semigroup, the Location theorem can be improved as follows:

Theorem 2.1 (Location theorem for finite semigroups) *Let J be a $\mathcal{J}$-class of a finite semigroup S, and let s and t be elements of J. The following conditions are equivalent:*

(1) $st \in J$,

(2) $st \in R(s) \cap L(t)$,

(3) $R(t) \cap L(s)$ *contains an idempotent,*
(4) $\bar{s}st = t$ *and* $st\bar{t} = s$ *for some inverse* $\bar{s}$ *of* s *and some inverse* $\bar{t}$ *of* t.

Proof. The equivalence of (1) and (2) follows from Theorem 1.8. The other equivalences follow from the Location theorem (Theorem 1.10). □

Let us mention a useful consequence.

Proposition 2.2 *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 1.8 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 2.1 shows that $H(t)$ contains an idempotent. Thus by Proposition 1.12, $H(t)$ 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 1.2. Thus $se = s = st$, whence $e = t$. □

The case $t = s$ is worth a separate statement, that should be compared with Proposition 1.12.

Corollary 2.3 *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 our study of finite semigroups by two results on maximal $\mathcal{J}$-classes.

Proposition 2.4 *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 1.4 whence $e = 1$ by Theorem 1.8. It follows that J contains a unique idempotent and by Proposition 1.11, a unique $\mathcal{R}$-class and a unique $\mathcal{L}$-class. Thus J is an $\mathcal{H}$-class and thus a group. □

Proposition 2.5 *Let* J *be a* $\leqslant_{\mathcal{J}}$*-maximal* $\mathcal{J}$*-class of a semigroup* S*. If* J *is finite, it is either regular or reduced to a single null element.*

Proof. Let J be a maximal $\mathcal{J}$-class. Suppose that J is not regular and contains two distinct element s and t. Then $s = utv$ and $t = xsy$ for some $x, y, u, v \in S^1$. Thus $s = uxsvy$ 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 $s \leqslant_{\mathcal{J}} ux$, it follows that $s \mathcal{J} ux$ since J is maximal. Similarly, $s \leqslant_{\mathcal{J}} svy$, whence $svy \in J$. Thus J contains ux, svy and their product. Therefore it is regular by Corollary 3.7. □

3 Inverses, weak inverses and regular elements

In this section, we study in more detail the notion of semigroup inverse introduced in Chapter II.

3.1 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 denote the set of all weak inverses (resp. inverses) of the element s by $W(s)$ (resp. $V(s)$).

Proposition 3.1 *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 $\bar{s}s\bar{s}s = \bar{s}s$. Furthermore, 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 3.2 *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 3.1. □

The case where $\bar{s}$ is a weak inverse (resp. an inverse) of s is depicted in Figure 3.1. 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.

$\bar{s}$	$*\ \bar{s}s$
$*\ s\bar{s}$	$s\bar{s}s$

$\bar{s}$	$*\ \bar{s}s$
$*\ s\bar{s}$	s

Figure 3.1. 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 3.3 *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 3.2, 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 1.2. 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 3.2 shows that $s\bar{s}_1$ and $s\bar{s}_2$ are idempotents of the same $\mathcal{H}$-class and hence are equal. It follows that $s\bar{s}_1 s = s\bar{s}_2 s$, 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 characterized as follows:

Proposition 3.4 *Let e and f be two idempotents of a semigroup S. Then e and f are conjugate if and only if they are $\mathcal{D}$-equivalent.*

Proof. 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 $e = uv \mathcal{D} vu = f$.

Conversely 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 $\bar{s}s\bar{s} = f\bar{s} = \bar{s}$. □

We conclude this section by an elementary result on idempotents.

Proposition 3.5 *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 $(ye)(ex)(ye) = yexye = ye$. □

3.2 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 (resp. $\mathcal{L}$-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 characterizations of regular elements.

Proposition 3.6 *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 3.1 shows that (2) implies (1), (3), (4) and (5). The equivalence of (3), (4) and (5) follows from Proposition 1.11.

(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 $s\bar{s}s = ststs = ees = e$ and $\bar{s}s\bar{s} = tststst = \bar{s}$. Thus s is regular. □

It is useful to restate Proposition 3.6 in terms of $\mathcal{D}$-classes.

Corollary 3.7 *Let 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 3.7 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 3.8 *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 the $\mathcal{K}$-class of each idempotent is trivial.*

Proof. We prove that if the $\mathcal{K}$-class of each idempotent is trivial, then S is $\mathcal{K}$-trivial.

(a) $\mathcal{K} = \mathcal{R}$. Suppose $a \mathcal{R} b$. Then there exist $c, d \in S^1$ such that $ac = b$, $bd = a$, whence $acd = a$. Thus $(1, cd) \in \mathrm{Stab}(a)$ and since S is finite, there exists a weak inverse $(\bar{v}, \bar{u})$ of $(1, cd)$ such that $\bar{v}a\bar{u} = a$. Since $\bar{u}cd\bar{u} = \bar{u}$, one has $\bar{u} \mathcal{R} \bar{u}c \mathcal{R} \bar{u}cd$, whence $\bar{u} = \bar{u}c = \bar{u}cd$ since $\bar{u}cd$ is idempotent. Therefore $a = \bar{v}a\bar{u} = \bar{v}a\bar{u}c = ac = b$ and hence 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) $\mathcal{K} = \mathcal{H}$. □

The case of the $\mathcal{H}$-relation, omitted in Proposition 3.8, is related to the study of aperiodic semigroups: a finite semigroup S is *aperiodic* if, for every $x \in S$, there exists an integer n such that $x^n = x^{n+1}$.

Proposition 3.9 *Let S be a finite semigroup. The following conditions are equivalent:*

(1) *S is aperiodic,*

(2) *S is $\mathcal{H}$-trivial,*

(3) *the groups in S are trivial.*

Proof. (1) implies (3). Let G be a group in S, with identity e, and let $x \in G$. By (1), there exists n such that $x^n = x^{n+1}$. Since x^n and x^{n+1} are both in G, this implies $x = e$. Thus G is trivial.

(3) implies (2). 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 $u^n a y^n = a$ for every n. Let us choose n such that u^n is idempotent. Since $u^n \mathcal{H} u^{n+1}$, $u^n = u^{n+1}$ and thus $a = u^n a y^n = u^{n+1} a y^n = u(u^n a y^n) = ua = b$. Therefore S is $\mathcal{H}$-trivial.

(3) implies (2) follows from the fact that each group in S is contained in an $\mathcal{H}$-class.

(2) implies (1). Let $x \in S$. Since S is finite, some power x^n of x belongs to a group. By (3), this group is trivial and x^n is idempotent. Furthermore, x^{n+1} belongs to the $\mathcal{H}$-class of x^n, which is a trivial group. Thus $x^n = x^{n+1}$. □

We conclude this section by another property of finite semigroups.

Proposition 3.10 *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 $e \mathcal{R}_T us \mathcal{L}_T s \mathcal{R}_T sv \mathcal{L}_T e$.*

$^*e$		us
sv		s

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 $e \in T$. □

4 Finite 0-simple semigroups

The fine structure of regular $\mathcal{D}$-classes will be detailed in Subsection 6. It relies on the notions of simple and 0-simple semigroups which form the subject of this section.

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. By Proposition II.2.8, a simple semigroup has a single $\mathcal{J}$-class, and a 0-simple semigroup has a single nonzero $\mathcal{J}$-class.

If a semigroup S has a zero 0, it is the unique minimal idempotent. An idempotent is called 0-*minimal* if, for every idempotent f, $f \leqslant e$ implies $f = e$ or $f = 0$.

4.1 Structure of 0-simple semigroups

The structure of 0-simple semigroups has been elucidated by Rees and is summarized in Theorem 5.4. The proof of this theorem is nontrivial and relies on a series of propositions and lemmas. Let us start with an elementary observation.

Lemma 4.1 *Let S be a 0-simple semigroup and let $s, t \in S$. If $sSt = 0$, then $s = 0$ and $t = 0$.*

Proof. By Lemma II.2.7, $S^2 = S$. If $s \neq 0$ and $t \neq 0$, then $SsS = StS = S$. Thus $SsStS = SsSStS = S^2 = S$ and $sSt \neq 0$. □

We now identify the $\mathcal{R}$-class of each 0-minimal idempotent.

Proposition 4.2 *Let S be a 0-simple semigroup. If e is a 0-minimal idempotent of S, then $R(e) \cup 0 = eS$.*

Proof. Let $R = R(e)$. By definition $R \subseteq eS$ and hence $R \cup 0 \subseteq eS$. Let s be a nonzero element of eS. Then $s \leqslant_{\mathcal{R}} e$ and thus $es = s$. Furthermore, since $SsS = S$, $e = usv$ for some $u, v \in S$. Thus $e = uesv = (ue)(sv)$ and by Proposition 3.5, eue and sve are mutually inverse elements. In particular, $f = (sve)(eue)$ is an idempotent such that $f \leqslant e$. Since e is 0-minimal, one has $e = f$, whence $s \mathcal{R} e$ and $s \in R$. Thus $eS \subseteq R \cup 0$. □

Proposition 4.3 *If R is a nonzero $\mathcal{R}$-class of a 0-simple semigroup, then $R \cup 0$ is a 0-minimal right ideal.*

Let us first show that $R \cup 0$ is a right ideal. Let $r \in R$ and let e be a 0-minimal idempotent. Since $S = SeS$, $r = uev$ for some $u, v \in S$. Thus $ev \neq 0$ and by Proposition 4.2, $ev \mathcal{R} e$. Therefore $R = R(uev) = uR(ev) = uR(e)$. It follows that $R \cup 0 = uR(e) \cup 0 = u(R(e) \cup 0) = ueS$.

Let us verify that $R \cup 0$ is 0-minimal. Let R' be a nonempty, nonzero right ideal contained in $R \cup 0$. Taking a nonzero element $r' \in R'$ and any element $r \in R$, we have $r' \mathcal{R} r$, whence $r \in r'S^1$ and $r \in R'S^1 = R'$. This shows that $R' = R \cup 0$. □

Proposition 4.2 can now be generalised as follows:

Proposition 4.4 *Let S be a 0-simple semigroup. For every $s \in S$, $R(s) \cup 0 = sS = sS^1$ and this right ideal is 0-minimal.*

Proof. The result is trivial if $s = 0$. If $s \neq 0$, Proposition 4.3 shows that $I = R(s) \cup 0$ is a 0-minimal right ideal containing s. Thus $sSS^1 = sS \subseteq sS^1 \subseteq I$. It follows that sS is a nonzero right ideal contained in I, and is consequently

equal to I. Thus $R(s) \cup 0 = sS = sS^1$. The 0-minimality of I follows from Proposition 4.3. □

A dual result holds of course for $L(s)$. This gives more insight on the Green's relations in S:

Corollary 4.5 *Let S be a 0-simple semigroup. If s and t are elements of S, then either $st = 0$ or $s \mathcal{R}\ st \mathcal{L}\ t$.*

Proof. If $st \neq 0$, then both s are t are nonzero elements. Furthermore stS^1 is contained in sS^1, and since sS^1 is a 0-minimal right ideal by Proposition 4.4, $stS^1 = sS^1$, and hence $s \mathcal{R}\ st$. Dually $st \mathcal{L}\ t$. □

Proposition 4.6 *A 0-simple semigroup contains a single nonzero $\mathcal{D}$-class and this $\mathcal{D}$-class is regular.*

Proof. Let $s, t \neq 0$. Then $sSt \neq 0$ by Lemma 4.1. Let $r \in sSt \setminus 0$. Since $r \in sS \cap St$, one has $r \in R(s) \cap L(t)$ by Proposition 4.4. Thus $s \mathcal{R}\ r \mathcal{L}\ t$ and $s \mathcal{D}\ t$. It follows that $S \setminus 0$ is a $\mathcal{D}$-class. Since $S \setminus 0$ contains an idempotent, this $\mathcal{D}$-class is regular by Corollary 3.7. □

The counterpart of Proposition 4.6 for simple semigroups can now be stated.

Proposition 4.7 *A simple semigroup contains a single $\mathcal{D}$-class. This $\mathcal{D}$-class is regular and each of its $\mathcal{H}$-classes is a group.*

Proof. Let S be a simple semigroup. Then S^0 is 0-simple and by Proposition 4.6, S is a regular $\mathcal{D}$-class of S^0. Furthermore, if $s \in S$, Corollary 4.5 shows that $s \mathcal{H}\ s^2$. It follows by Proposition 1.12 that $H(s)$ is a group. □

5 Rees matrix semigroups

Rees matrix semigroups play an important role in semigroup theory. 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

$$(i, g, j)(i', g', j') = (i, gp_{j,i'}g', j') \tag{5.1}$$

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 denote by $M^0(G, I, J, P)$ the semigroup, called a *Rees matrix semigroup with zero*, defined on the set $(I \times G \times J) \cup 0$ by the operation

$$(i, g, j)(i', g', j') = \begin{cases} (i, gp_{j,i'}g', j') & \text{if } p_{j,i'} \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

As suggested by the terminology, Rees matrix semigroups can be equivalently defined as semigroups of matrices. We first define an addition on G^0 by setting $g + g' = 0$ for all $g \in G^0$. Then, for all $g, g', h \in G^0$,

$$(g + g')h = gh + g'h = 0 \text{ and } h(g + g') = hg + hg' = 0,$$

whereby G^0 is equipped with a structure of idempotent semiring. We now identify each element of $M^0(G, I, J, P)$ with an $I \times J$ matrix with entries in G^0: 0 is identified with the null matrix and (i, g, j) with the matrix whose sole nonzero entry is g in row i and column j. The product of two matrices X and Y in $M^0(G, I, J, P)$ is XPY. Note that all products can be calculated using only trivial sums $0 + g = g + 0 = g$.

Example 5.1 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.1.5.

The Rees matrix semigroups with zero that arise in the study of 0-simple semigroups are all regular. This property depends only on the sandwich matrix.

Proposition 5.1 *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 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 $p_{j',i} \neq 0$. Consequently, every row and every 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

$$\begin{aligned} s\bar{s}s &= (i, gp_{j,i_j}p_{j,i_j}^{-1}g^{-1}p_{j_i,i}^{-1}p_{j_i,i}g, j) = s \text{ and} \\ \bar{s}s\bar{s} &= (i_j, p_{j,i_j}^{-1}g^{-1}p_{j_i,i}^{-1}p_{j_i,i}gp_{j,i_j}p_{j,i_j}^{-1}g^{-1}p_{j_i,i}^{-1}, j_i) = s' \end{aligned}$$

Thus S is regular. □

Green's relations in a regular Rees matrix semigroup with zero are easy to describe.

Proposition 5.2 *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 $S \setminus 0$ are in the same $\mathcal{D}$-class. Furthermore, if $s = (i, g, j)$ and $s' = (i', g', j')$ are two elements of $S \setminus 0$, then*

$$s \leqslant_{\mathcal{R}} s' \iff s \mathcal{R} s' \iff i = i', \tag{5.2}$$
$$s \leqslant_{\mathcal{L}} s' \iff s \mathcal{L} s' \iff j = j', \tag{5.3}$$
$$s \leqslant_{\mathcal{H}} s' \iff s \mathcal{H} s' \iff i = i' \text{ and } j = j'. \tag{5.4}$$

Proof. Proposition 5.1 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 5.1 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 (5.2). Property (5.3) is dual and (5.7) is the conjunction of (5.2) and (5.6).

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 $S \setminus 0$. Finally, if e and f are nonzero idempotents such that $e \leqslant f$, $e \mathcal{H} f$ by (3), and hence $e = f$ by Proposition 1.12. 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 5.1. The product of two elements when $p_{j,i'} \neq 0$.

A slightly more precise result holds for Rees matrix semigroups.

Proposition 5.3 *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*

$$s \leqslant_{\mathcal{R}} s' \iff s \mathcal{R} s' \iff i = i', \tag{5.5}$$
$$s \leqslant_{\mathcal{L}} s' \iff s \mathcal{L} s' \iff j = j', \tag{5.6}$$
$$s \leqslant_{\mathcal{H}} s' \iff s \mathcal{H} s' \iff i = i' \text{ and } j = j'. \tag{5.7}$$

Proof. The proposition mostly follows from Proposition 5.2 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 1.12, $H(s)$ is a group. Consequently, each $\mathcal{H}$-class is a group. □

We can now state the main theorem of this section.

Theorem 5.4 (Rees-Sushkevich theorem)

(1) *A semigroup is simple if and only if it is isomorphic to some Rees matrix semigroup.*

(2) *A semigroup is 0-simple if and only if it is isomorphic to some regular Rees matrix semigroup with zero.*

Proof. By Proposition 5.2, every Rees matrix semigroup with zero is 0-simple. Similarly by Proposition 5.3, every Rees matrix semigroup is simple.

Let S be a 0-simple semigroup. Let $(R_i)_{i\in I}$ (resp. $(L_j)_{j\in J}$) be the set of $\mathcal{R}$-classes (resp. $\mathcal{L}$-classes) of S and let e be an idempotent of S. We denote by $H_{i,j}$ the $\mathcal{H}$-class $R_i \cap L_j$. By Propositions 1.12 and 4.6, 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 1.2, $er_j = r_j$ and $s_ie = s_i$. Consequently, by Green's lemma the map $g \to s_igr_j$ from G into $H_{i,j}$ is a bijection. It follows that each element of $S \setminus 0$ admits a unique representation of the from s_igr_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 $p_{j,i} = r_js_i$. By Theorem 1.10, $r_js_i \in G$ if $H_{i,j}$ contains an idempotent and $r_js_i = 0$ otherwise. Define a map $\varphi : S \to M^0(G, I, J, P)$ by setting

$$\varphi(s) = \begin{cases} (i, g, j) & \text{if } s = s_igr_j \\ 0 & \text{if } s = 0 \end{cases}$$

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

$$\varphi(s)\varphi(s') = (i, g, j)(i', g', j') = \begin{cases} (i, gr_js_ig', j) & \text{if } H_{i',j} \text{ contains an idempotent} \\ 0 & \text{otherwise} \end{cases}$$

Since $s \in H_{i,j}$ and $s' \in H_{i',j'}$, Theorem 2.1 shows that $ss' \neq 0$ if and only if $H_{i',j}$ contains an idempotent and in this case, $ss' = s_igr_js_ig'r_j = s_i(gr_js_ig')r_j$. Therefore φ is a morphism, bijective by construction and hence is an isomorphism.

The case of simple semigroups can be handled in a similar way. □

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.

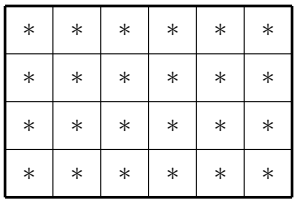


Figure 5.2. The rectangular band $B(4,6)$.

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.

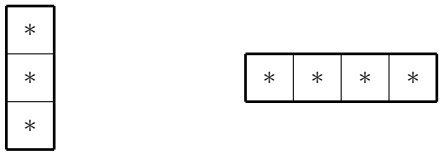


Figure 5.3. A left zero semigroup and a right zero semigroup.

6 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 6.1 *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, $L(s)$ (resp. $L(r)$) contains an idempotent e (resp. f). Thus $se = s$ and $rf = r$ by Proposition 1.2. It follows that $r = s(eu)$ and $s = r(fv)$. Furthermore, $eu \mathcal{L} su$ since $e \mathcal{L} s$ and thus $su \in D$. Similarly, $fv \in D$ and hence $s \mathcal{R} r$ in D^0. Dually, $r \mathcal{R} t$ 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. □

6.1 Structure of the minimal ideal

Proposition 6.2 *Let S be a finite semigroup. Then S has a unique minimal ideal. This ideal is a 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 1.12 that $H(s)$ 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 simple. □

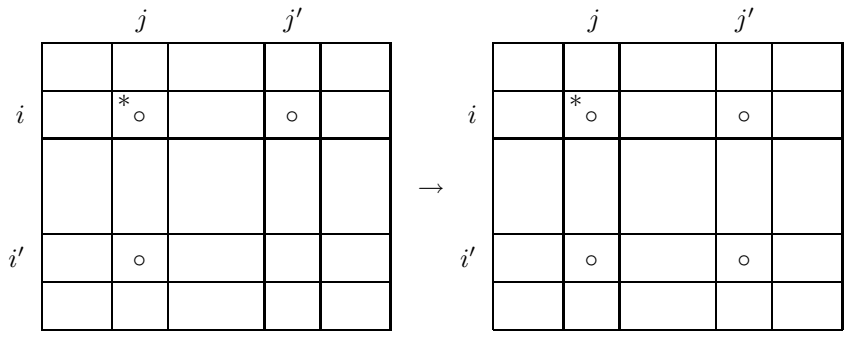
6.2 Blocks

A 0-simple semigroup is called a *block* if it is generated by its groups. The term "block" refers to a result of R. L. Graham [9] on 0-simple semigroups.

Theorem 6.3 *Let S^0 be a 0-simple semigroup and let T be the subsemigroup of S^0 generated by the union of all groups in S. Let $(B_i)_{i\in I}$ be the family of all regular $\mathcal{D}$-classes of T. The following properties hold:*

(1) *each B_i is a block,*
(2) *for all $i \neq j$, $B_iB_j = B_jB_i = 0$,*
(3) *for all $s \in S \setminus \bigcup_{i\in I} B_i$, $s^2 = 0$.*

The B_i's are called the *blocks* of S^0. It is easy to compute the blocks given the egg-box picture of S. You are given two different kind of tokens, stars and circles. First put a star and a circle in each $\mathcal{H}$-class which is a group (that is, containing an idempotent). Next, play the following game, as long as you can: each time there is a star in $H_{i,j}$ and circles in $H_{i,j'}$ and $H_{i',j}$, add a circle in $H_{i',j'}$.



When the game is over, one gets the block structure of S.

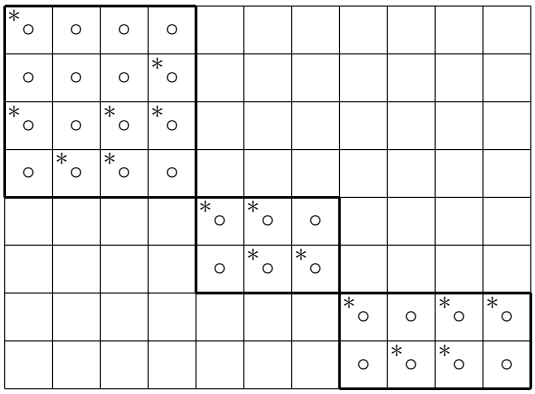


Figure 6.1. The block structure of a 0-simple semigroup.

Normal form, isomorphism problem. TO DO.

6.3 Examples

Rectangular bands, right groups. TO DO.

7 Green's relations in subsemigroups and quotients

Let us start with a trivial observation: Green's relations are stable under morphisms.

Proposition 7.1 *Let $\varphi : S \to T$ be a surjective 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 (resp. 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 denote by $\mathcal{K}_S$ (resp. $\mathcal{K}_T$) the Green's relation or preorder defined in the semigroup S (resp. T).

7.1 Green's relations in subsemigroups

We first consider the case of subsemigroups.

Proposition 7.2 *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 1.2 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 7.2 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 $T = E(S) = \{a^2, b^2, 0\}$. 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 7.3 *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, a similar result holds without any regularity assumption.

Proposition 7.4 *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 $t = ete$, $s = ese = eteu = eteeue$. Thus $s \leqslant_{\mathcal{R}_T} t$. The proof for the other relations is similar. □

A useful consequence of Proposition 7.2 is the following corollary:

Corollary 7.5 *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 7.2. □

7.2 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 7.6 *Let $\mathcal{K}$ be one of the relations $\mathcal{R}$, $\mathcal{L}$, $\mathcal{H}$ 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 7.1. □

More precise results hold for finite semigroups.

Proposition 7.7 *Suppose that S is finite. Let J be a $\mathcal{J}$-class of T and let I be a 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 (resp. $\mathcal{L}$-class) of S contained in I maps under φ onto an $\mathcal{R}$-class (resp. $\mathcal{L}$-class) 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. TO DO. Arbib p. 160. □

The following example shows that Proposition 7.7 does not extend to $\mathcal{H}$.

Example 7.1 Let M and N be the monoids generated by the following partial transformations:

M	1	2	3	4	5	6
a	2	1	5	6	3	4
b	3	4	0	0	0	0

N	1	2	3	4
a	2	1	4	3
b	3	4	0	0

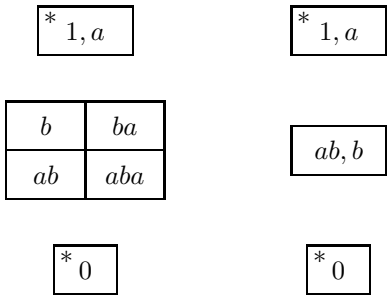
Their elements are respectively

M	1	2	3	4	5	6
a	2	1	5	6	3	4
b	3	4	0	0	0	0
ab	4	3	0	0	0	0
ba	5	6	0	0	0	0
bb	0	0	0	0	0	0
aba	6	5	0	0	0	0

N	1	2	3	4
a	2	1	4	3
b	3	4	0	0
ab	4	3	0	0
bb	0	0	0	0

Thus the monoid $M = \{1, a, b, ab, ba, bab, 0\}$ is presented on A by the relations $aa = 1$, $aab = 0$ and $bb = 0$. The monoid $N = \{1, a, b, ab, 0\}$ is presented on A by the relations $aa = 1$, $ba = ab$ and $bb = 0$.

The $\mathcal{J}$-class structures 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.

However, the following result holds

Proposition 7.8 *Suppose that S is finite. For each group H in T there exists a group G in S such that $\varphi(G) = H$.*

Proof. TO DO. □

8 Green's relations in $\mathfrak{T}(E)$.

Given an element $a \in \mathfrak{T}(E)$, 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

$$p \sim_a q \Longleftrightarrow p \cdot a = q \cdot a$$

Finally, we set $\mathrm{rank}(a) = \mathrm{Card}(\mathrm{Im}(a)) = \mathrm{Card}(\mathrm{Ker}(a))$. For example, if

$$a = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 1 & 4 & 5 & 5 & 5 & 4 & 1 \end{pmatrix}$$

we have $\mathrm{Im}(a) = \{1, 4, 5\}$ and $\mathrm{Ker}(a) = 17/26/345$.

Lemma 8.1 *Let $a, b \in \mathfrak{T}(E)$. 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 $\mathrm{Ker}(ab) \subseteq \mathrm{Ker}(a)$. □

Proposition 8.2 *Let a, b be elements of $\mathfrak{T}(E)$. 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 $u \in \mathfrak{T}(E)$, 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 $u \in \mathfrak{T}(E)$, such that $au = b$ 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 $u, v \in \mathfrak{T}(E)$ 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 $\mathrm{Card}(\mathrm{Im}(a)) = \mathrm{Card}(\mathrm{Ker}(a)) \leqslant \mathrm{Card}(\mathrm{Im}(b))$. 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. □

$*\,1$	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 :

$$
\begin{array}{cccc}
c^2 = 1 & a^2b = a^3 & a^2c = b^2 & b^3 = b^2a \\
b^2c = a^2 & ca^2 = b^2 & cb^2 = a^2 & a^4 = 0 \\
aba^2 = ab^2 & abac = abab & ab^2a = a^3 & abca = a^2 \\
abcb = ab & acab = abab & acba = a^3 & ba^3 = b^2a \\
bab^2 = ba^2 & babc = baba & baca = ba & bacb = b^2 \\
b^2a^2 = 0 & b^2ab = 0 & b^2ac = ba^2 & bcab = b^2a \\
bcba = baba & caba = baba & cab^2 = ba^2 & cba^2 = ab^2 \\
cbab = abab & ababa = aba & acaca = aca & acacb = acb \\
acbcb = acbca & babab = bab & bcaca = acbca & bcacb = acbca \\
bcbca = bca & bcbcb = bcb & &
\end{array}
$$

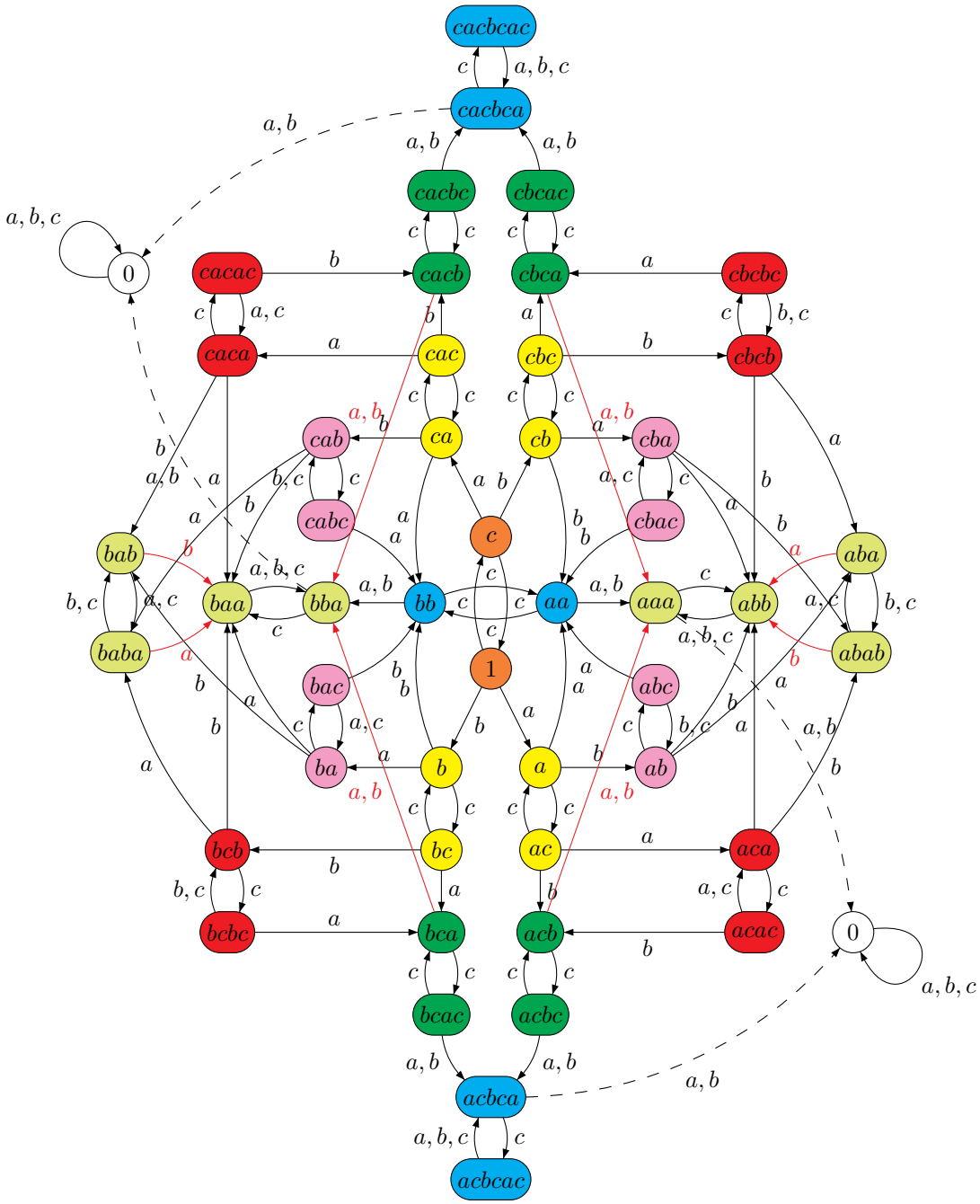


Figure 8.1. The right Cayley graph of S. To avoid unesthetic crossing lines, the zero is represented twice in this diagram.

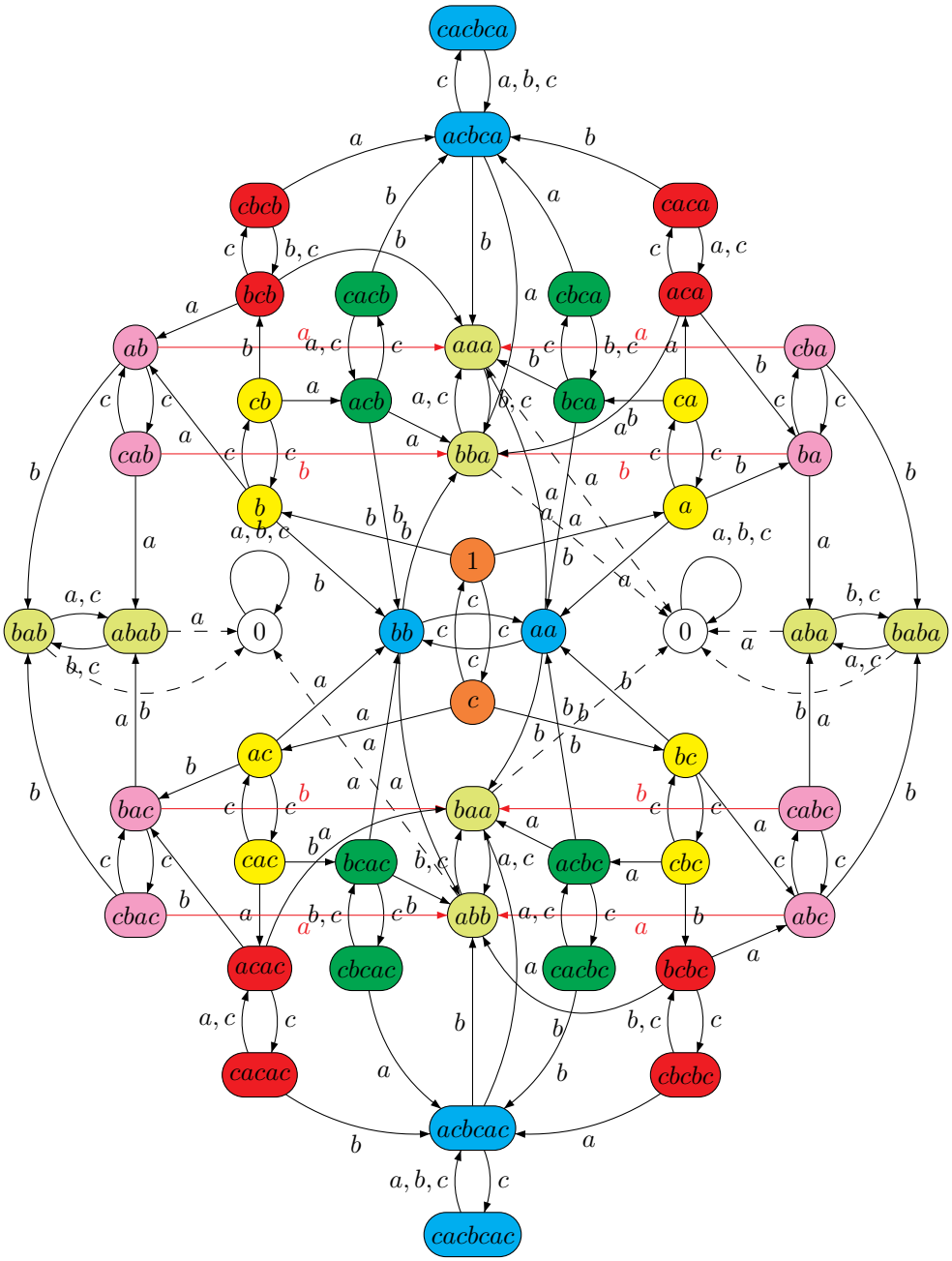


Figure 8.2. The left Cayley graph of S. To avoid unesthetic crossing lines, the zero is represented twice in this diagram.

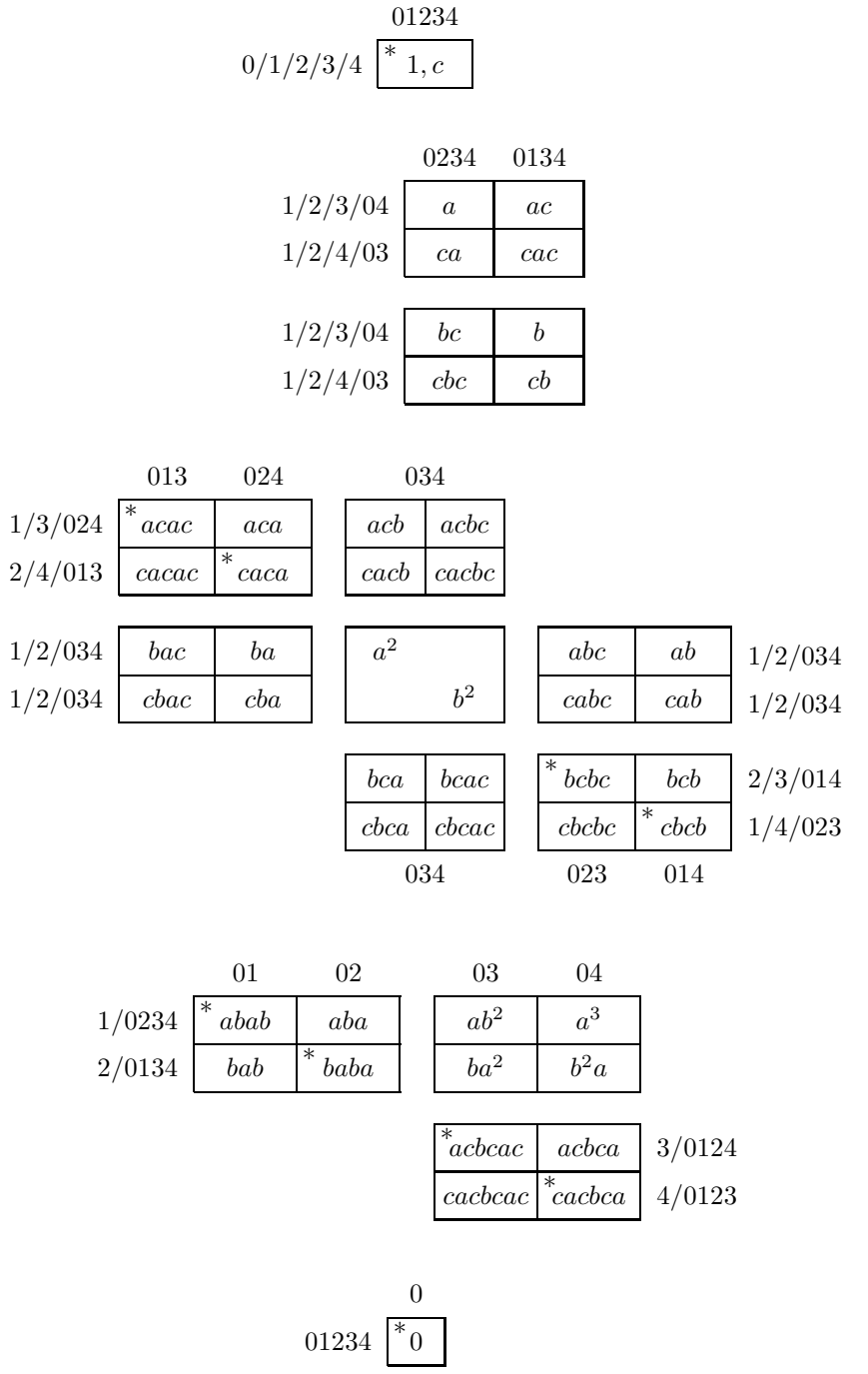


Figure 8.3. The $\mathcal{D}$-class structure of S.

The idempotents of S are 1, $a^4 = 0$ and

$$\begin{array}{llll} e_1 = acac & e_2 = cbcb & e_3 = caca & e_4 = bcbc \\ e_5 = abab & e_6 = cacbca & e_7 = baba & e_8 = acbcac \end{array}$$

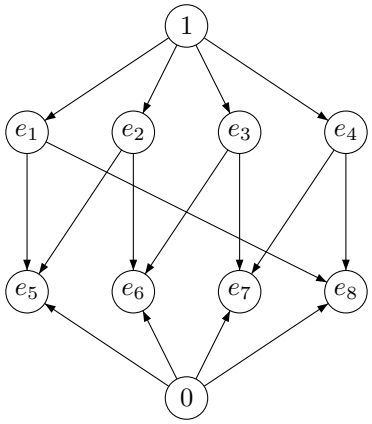


Figure 8.4. The lattice of idempotents.

9 Exercises.

Exercice 1 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 $\bar{s}s = \bar{t}t$.

Exercice 2 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 $s = t$.

Exercice 3 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.

Exercice 4 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 $\mathcal{I}(M)$.

Chapter VI

Varieties

1 Birkhoff varieties

A *Birkhoff 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 family of semigroups of $\mathbf{V}$, the product $\prod_{i \in I} S_i$ is also in $\mathbf{V}$.

The semigroups of a Birkhoff variety can be finite or infinite.

The definition of a Birkhoff variety can be readily generalized to monoids, groups, ordered semigroups, ordered monoids, etc.

Example 1.1

(1) The class of all semigroups forms a Birkhoff variety.

(2) The smallest Birkhoff variety is the trivial variety, consisting only of the empty semigroup and of the semigroup 1.

(3) The class of all commutative semigroups forms a Birkhoff variety.

(4) The class of all groups does not form a Birkhoff variety of monoids. Indeed $\mathbb{Z}$ is a group, but $\mathbb{N}$ is a submonoid of $\mathbb{Z}$ which is not a group.

Let $\mathbf{V}$ be a Birkhoff variety of semigroups and let A be an alphabet. Consider the collection $\mathcal{F}(A, \mathbf{V})$ of all morphisms $\varphi : A^+ \to S$, where S is a semigroup of $\mathbf{V}$. For each such morphism φ, denote by $\sim_\varphi$ the nuclear congruence of φ, and let $\sim$ be the intersection of all these congruences when φ ranges over $\mathcal{F}(A, \mathbf{V})$. Finally, set

$$\mathbf{F}_\mathbf{V}(A) = A^+/\sim$$

and let $\pi : A^+ \to \mathbf{F}_\mathbf{V}(A)$ be the natural morphism. As a subsemigroup of a product of semigroups of $\mathbf{V}$, this semigroup belongs to $\mathbf{V}$. It is called the $\mathbf{V}$-*free semigroup* on A. This terminology is justified by the following universal property, which relativizes to $\mathbf{V}$ the universal property of A^+ given by Proposition II.4.1.

Proposition 1.1 *If φ is a function from A into a semigroup S of $\mathbf{V}$, there exists a unique semigroup morphism $\bar{\varphi} : F_\mathbf{V}(A) \to S$ such that, for each $a \in A$, $\varphi(a) = \bar{\varphi}(\pi(a))$. Moreover, $\bar{\varphi}$ is surjective if and only if the set $\varphi(A)$ generates S.*

Proof. First, φ can be extended (in a unique way) into a morphism from A^+ into S. Now, since the nuclear congruence of φ is coarser than $\sim$, there exists by Proposition II.2.11 a unique morphism $\bar{\varphi} : F_{\mathbf{V}}(A) \to S$ such that $\varphi = \bar{\varphi} \circ \pi$.

Clearly, if $\bar{\varphi}$ is surjective, $\varphi : A^+ \to S$ is onto and thus A generates S. Conversely, if $\varphi(A)$ generates S, φ is onto and $\bar{\varphi}$ is also onto. □

Corollary 1.2 *An A-generated semigroup belongs to $\mathbf{V}$ if and only if it is a quotient of $F_{\mathbf{V}}(A)$.*

Proof. Let S be an A-generated semigroup of $\mathbf{V}$. By definition, there exists a surjective morphism $\varphi : A^+ \to S$. In particular, $\varphi \in \mathcal{F}(A)$ and thus $\pi(u) = \pi(v)$ implies $\varphi(u) = \varphi(v)$. By Proposition II.2.11, S is a quotient of $F_{\mathbf{V}}(A)$.

Conversely, assume that S is a quotient of $F_{\mathbf{V}}(A)$. Since $F_{\mathbf{V}}(A)$ belongs to $\mathbf{V}$, S is also in $\mathbf{V}$. □

Example 1.2 Let $\mathbf{V}$ be the Birkhoff variety of commutative monoids. It is easy to see that the $\mathbf{V}$-free monoid (called, in this case, the free commutative monoid), is the additive monoid $\mathbb{N}^A$, where the addition is defined componentwise: $(n_a)_{a\in A} + (n'_a)_{a\in A} = (n_a + n'_a)_{a\in A}$.

A convenient way to define varieties is to use identities. Let A be an alphabet and let $u, v \in A^+$. A semigroup S *satisfies the identity* $u = v$ if and only if, for each morphism of semigroups $\varphi : A^+ \to S$, $\varphi(u) = \varphi(v)$. For example, a semigroup satisfies the identity $xyx = x$ if, for each $x, y \in S$, $xyx = x$. A semigroup is commutative if and only if it satisfies the identity $xy = yx$; it is idempotent if and only if it satisfies the identity $x = x^2$.

A Birkhoff variety $\mathbf{V}$ satisfies a given identity if every semigroup 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 $\mathbf{V}$-free semigroups.

Proposition 1.3 *Let $\pi_{\mathbf{V}}$ be the natural morphism from A^+ onto the $\mathbf{V}$-free semigroup on A. Given two words u and v of 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 $\pi_{\mathbf{V}}(u) = \pi_{\mathbf{V}}(v)$ since $F_{\mathbf{V}}(A) \in \mathbf{V}$. Conversely if $\pi_{\mathbf{V}}(u) = \pi_{\mathbf{V}}(v)$, then by definition $\varphi(u) = \varphi(v)$ for every $\varphi \in \mathcal{F}(A, \mathbf{V})$, and thus $u = v$ is an identity of $\mathbf{V}$. □

Corollary 1.4 *Let $\mathbf{V}$ and $\mathbf{W}$ be two varieties satisfying the same identities on the alphabet A. Then the $\mathbf{V}$-free semigroup and the $\mathbf{W}$-free semigroup on A are isomorphic.*

Proof. Indeed, by Proposition 1.3, the nuclear congruences of $\pi_{\mathbf{V}}$ and $\pi_{\mathbf{W}}$ are the same. Therefore, the semigroups $\pi_{\mathbf{V}}(A^+)$ and $\pi_{\mathbf{W}}(A^+)$ are equal. □

Let E be a set of identities. The class $[\![E]\!]$ of semigroups is defined to be the class of semigroups satisfying the identities of E. The next result explains the importance of identities.

Theorem 1.5 (Birkhoff's theorem) *A class of semigroups is a Birkhoff variety if and only if it can be defined by a set of identities.*

Proof. We first prove that every class defined by a set of identities is a Birkhoff variety. Since Birkhoff varieties are closed under intersection, it suffices to establish the result when the class is defined by a single identity, say $u = v$. Let S be an ordered semigroup satisfying this identity. Then clearly, every subsemigroup of S satisfies the same identity. Furthermore, if $\pi : S \to T$ is an onto morphism, T also satisfies $u = v$. Indeed, if $\varphi : A^+ \to T$ is a morphism, there exists by Corollary II.4.3 a morphism $\psi : A^+ \to S$ such that $\varphi = \pi \circ \psi$. Now, $\psi(u) = \psi(v)$ since S satisfies the identity $u = v$ and thus $\pi(\psi(u)) = \pi(\psi(v))$. Therefore, T satisfies the identity $u = v$. Finally, if $(S_i)_{i\in I}$ is a family of semigroups satisfying the identity $u = v$, their product $S = \prod_{i\in I} S_i$ also satisfies this identity. Indeed, let π_i denotes the projection from S onto S_i and let φ be a morphism from A^+ into S. Since $\pi_i \circ \varphi$ is a morphism from A^+ into S_i, $\pi_i \circ \varphi(u) = \pi_i \circ \varphi(v)$. As this holds for each i, we have $\varphi(u) = \varphi(v)$.

Let now $\mathbf{V}$ be a Birkhoff variety. Let E be the class of all identities satisfied by $\mathbf{V}$ and let $\mathbf{W} = [\![E]\!]$. Clearly $\mathbf{V} \subseteq \mathbf{W}$. Let $S \in \mathbf{W}$ and let Σ be a generating set of S. Then there exists an onto morphism $\varphi : \Sigma^+ \to S$. Let $\pi : \hat{\Sigma}^+ \to F_{\mathbf{V}}(\Sigma)$ be the natural morphism. Let $u, v \in \Sigma^+$. By Proposition 1.3, if $\pi(u) = \pi(v)$ then $u = v$ is an identity of $\mathbf{V}$ and thus, is satisfied by S. In particular, $\varphi(u) = \varphi(v)$. It follows by Proposition II.2.11 that φ factors through π. Therefore S is a quotient of $F_{\mathbf{V}}(\Sigma)$ and by Corollary 1.2, S is in $\mathbf{V}$. Thus $\mathbf{V} = [\![E]\!]$. □

2 Varieties of finite semigroups

The semigroups occurring in automata theory are mostly finite, which motivates the following definition.

A *variety of finite semigroups* is a class of finite 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 $\mathbf{V}$.

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 2.1

(1) The class $\mathbf{S}$ of all finite semigroups forms a variety of finite semigroups.

(2) The smallest variety of finite semigroups is the trivial variety, consisting of the empty semigroup and of the semigroup with one element 1. This variety is denoted by $\mathbf{1}$.

(3) The class of all commutative finite semigroups forms a variety of finite semigroups, denoted by **Com**.

Let us give an important example of variety of finite monoids, which should be compared with Example 1.1 (4).

Proposition 2.1 *The class of all finite groups is a variety of finite monoids, denoted by* **G**.

Proof. Finite groups are lcosed under quotients ad finite direct products and it follows from Proposition II.5.6 that a submonoid of a finite group is a group. □

In the sequel, we shall use the term "variety" as a shorthand for variety of finite semigroups (resp. monoids, etc.).

The *supremum* of two varieties $\mathbf{V}_1$ and $\mathbf{V}_2$ is the smallest variety containing $\mathbf{V}_1$ and $\mathbf{V}_2$. It is denoted by $\mathbf{V}_1 \vee \mathbf{V}_2$.

Let $\mathcal{C}$ be a class of semigroups. The class of all semigroups that divide a finite product of semigroups of $\mathcal{C}$ is a variety, denoted $\langle \mathcal{C} \rangle$ and called the *variety generated* by $\mathcal{C}$. It is the smallest variety containing $\mathcal{C}$.

It is tempting to use identities to obtain a counterpart of Theorem 1.5 for varieties of finite algebras, but an incursion into topology is necessary in order to achieve this goal. This approach is very similar to the introduction of p-adic numbers in number theory and is a good illustration of the following quotation of Marshall Stone: 'A cardinal principle of modern mathematical research may be stated as a maxim: "One must always topologize"'.

3 Profinite algebras

The details are provided for semigroups, but the results can be readily adapted to monoids.

A *metric semigroup* is a semigroup S equipped with a metric d, such that (S, d) is a complete metric space and the multiplication of S is uniformly continuous. Morphisms between two metric semigroups are required to be uniformly continuous.

If X is a subset of S, the *metric subsemigroup* of S generated by X is by definition the smallest *closed* subsemigroup of S containing X.

Finite semigroups can be considered as metric semigroups, equipped with the discrete metric. More precisely, if S is a finite semigroup, the discrete metric d is defined by

$$d(s,t) = \begin{cases} 0 & \text{if } s = t \\ 1 & \text{otherwise} \end{cases}$$

In this section, we shall systematically consider finite semigroups as metric semigroups without any further warning.

Another important example of metric semigroup is the free pro-**V** semigroup on A, that we now define. A semigroup S *separates* two words u and v of the free semigroup A^+ if there exists a morphism φ from A^+ onto S such that $\varphi(u) \neq \varphi(v)$. Let **V** be a variety of finite semigroups. We set

$$r_{\mathbf{V}}(u,v) = \min\{\mathrm{Card}(S) \mid S \text{ is a semigroup of } \mathbf{V} \text{ that separates } u \text{ and } v\}$$

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 3.1 *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 semigroup of $\mathbf{V}$ 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 S be a semigroup of $\mathbf{V}$ separating u and w. Then S 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 $d_{\mathbf{V}}(u,w) \leqslant \max\{d_{\mathbf{V}}(u,v), d_{\mathbf{V}}(v,w)\}$. □

If $\mathbf{V}$ is the variety of all finite semigroups, we simplify the notation $d_{\mathbf{V}}$ to d.

Proposition 3.2 *The function d is an ultrametric on A^+.*

Proof. Suppose that $d(u,v) = 0$. In particular, the syntactic semigroup of $\{u\}$ does not separate u from v, showing that $u = v$. Thus by Proposition 3.1, d is an ultrametric. □

For the metric d, the closer are two words, the larger is the semigroup needed to separate them.

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 $\mathbf{V}$ is the variety of commutative semigroups, $d_{\mathbf{V}}(ab,ba) = 0$, since there is no way to separate ab and ba by a commutative semigroup. To work around this difficulty, we first observe that, by Proposition 3.1, 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 3.2 can be generalized as follows.

Proposition 3.3
(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 3.1, 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 A^+ for d, denoted by $\widehat{A^+}$, is called the *free profinite semigroup* on A. The completion of the metric space $(A^+/\!\sim_{\mathbf{V}}, d_{\mathbf{V}})$, denoted by $\hat{F}_{\mathbf{V}}(A)$, is called the *free pro-*$\mathbf{V}$ *semigroup* on A. It satisfies the following properties:

Proposition 3.4 *Let $\mathbf{V}$ be a variety of semigroups and A a finite alphabet.*
(1) *The semigroup $\hat{F}_{\mathbf{V}}(A)$ is compact.*
(2) *There is a surjective morphism $\pi_{\mathbf{V}}$ from $\widehat{A^+}$ onto $\hat{F}_{\mathbf{V}}(A)$.*

(3) *Every morphism from* $\widehat{A^+}$ *into a semigroup of* $\mathbf{V}$ *factorizes through* $\pi_{\mathbf{V}}$.
(4) *An* A*-generated finite semigroup belongs to* $\mathbf{V}$ *if and only if it is a quotient of* $\hat{F}_{\mathbf{V}}(A)$.

Proof. (1) Since $\hat{F}_{\mathbf{V}}(A)$ 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

$$u \sim_n v \text{ if and only if } \varphi(u) = \varphi(v) \text{ for every morphism } \varphi \text{ from } A^+ \text{ onto a semigroup of size } \leqslant n \text{ of } \mathbf{V}.$$

Since A is finite, there are only finitely many morphisms from A^+ onto a semigroup 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 semigroup of $\mathbf{V}$ 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) Let $\pi_{\mathbf{V}}$ be the natural morphism from A^+ onto $A^+/\sim_{\mathbf{V}}$. Since $d_{\mathbf{V}}(u, v) \leqslant d(u, v)$, $\pi_{\mathbf{V}}$ is uniformly continuous, and can be extended into a uniformly continuous morphism from $\widehat{A^+}$ onto $\hat{F}_{\mathbf{V}}(A)$.

(3) Let φ be a morphism from $\widehat{A^+}$ into a semigroup S of $\mathbf{V}$. Up to replacing S by $\varphi(S)$, we may assume that φ is onto. Since A^+ is dense in $\widehat{A^+}$, and S 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.2.11 shows that there is a surjective morphism π from $A^+/\sim_{\mathbf{V}}$ onto S such that $\varphi = \pi \circ \pi_{\mathbf{V}}$. We claim that this morphism is uniformly continuous. Indeed if $d_{\mathbf{V}}(u, v) < 2^{-|S|}$, then u and v cannot be separated by S, and hence $\varphi(u) = \varphi(v)$. Since A^+ is dense in $\hat{F}_{\mathbf{V}}(A)$, π can be extended by continuity into a surjective morphism from $\hat{F}_{\mathbf{V}}(A)$ onto S. Thus S is a quotient of $\hat{F}_{\mathbf{V}}(A)$.

(4) If S is an A-generated semigroup of $\mathbf{V}$, there exists a surjective morphism φ from A^+ onto S. Following the argument used in (3), if $d_{\mathbf{V}}(u, v) < 2^{-|S|}$, then $\varphi(u) = \varphi(v)$, and thus φ is uniformly continuous with respect to $d_{\mathbf{V}}$. Therefore φ can be extended into a uniformly continuous morphism from $\hat{F}_{\mathbf{V}}(A)$ onto S.

Conversely, assume that S is a finite quotient of $\hat{F}_{\mathbf{V}}(A)$ and let $\pi : \hat{F}_{\mathbf{V}}(A) \to S$ be a surjective morphism. The set

$$D = \{(u, v) \in \hat{F}_{\mathbf{V}}(A) \times \hat{F}_{\mathbf{V}}(A) \mid \pi(u) = \pi(v)\}$$

is the inverse image under π of the diagonal of $S \times S$, and since S is discrete and π is continuous, it is a clopen subset of $\hat{F}_{\mathbf{V}}(A) \times \hat{F}_{\mathbf{V}}(A)$. Let $\mathcal{F}$ be the class of all morphisms from $\hat{F}_{\mathbf{V}}(A)$ onto a semigroup of $\mathbf{V}$. For each $\varphi \in \mathcal{F}$, let

$$C_\varphi = \{(u, v) \in \hat{F}_{\mathbf{V}}(A) \times \hat{F}_{\mathbf{V}}(A) \mid \varphi(u) \neq \varphi(v)\}$$

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 semigroup 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 $D \cup (C_\varphi)_{\varphi \in \mathcal{F}}$ is a covering of $\hat{F}_{\mathbf{V}}(A) \times \hat{F}_{\mathbf{V}}(A)$ by open sets, and since $\hat{F}_{\mathbf{V}}(A)$ is compact, it admits a finite subcovering, say $D \cup (C_\varphi)_{\varphi \in F}$. Therefore, if $\varphi(u) = \varphi(v)$ for each $\varphi \in F$, then $\pi(u) = \pi(v)$. Consequently S is a quotient of a subsemigroup of the finite semigroup $\prod_{\varphi \in F} \varphi(\hat{F}_{\mathbf{V}}(A))$ and thus belongs to $\mathbf{V}$. □

We now extend the notion of identity as follows. Let A be a finite alphabet and let $u, v \in \widehat{A^+}$. A finite semigroup S *satisfies the identity* $u = v$ if and only if, for each morphism $\varphi : \widehat{A^+} \to S$, $\varphi(u) = \varphi(v)$.

A variety $\mathbf{V}$ satisfies a given identity if every semigroup 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-$\mathbf{V}$ semigroups.

Proposition 3.5 *Let A be a finite alphabet. Given two elements 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 semigroup 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 3.4, $\varphi(u) = \varphi(v)$ for every morphism φ from $\widehat{A^+}$ into a semigroup of $\mathbf{V}$, and thus $u = v$ is an identity of $\mathbf{V}$. □

Corollary 3.6 *Let $\mathbf{V}$ and $\mathbf{W}$ be two varieties of finite semigroups satisfying the same identities on the alphabet A. Then $\hat{F}_{\mathbf{V}}(A)$ and $\hat{F}_{\mathbf{W}}(A)$ are isomorphic.*

In particular, an identity of a semigroup of $\mathbf{V}$ can be given as a pair (u, v) of elements of $\hat{F}_{\mathbf{V}}(A)$. We are now ready to state the generalization of Theorem 1.5. Given a set E of identities, we denote by $[\![E]\!]$ the class of finite semigroups satisfying all the identities of E.

Theorem 3.7 (Reiterman's theorem) *A class of finite semigroups is a variety if and only if it can be defined by a set of identities.*

Proof. The fact that every class of finite semigroups defined by a set of identities is a variety can be proved exactly as in Theorem 1.5.

Let now $\mathbf{V}$ be a variety. Let E be the class of all identities which are satisfied by every semigroup of $\mathbf{V}$ and let $\mathbf{W} = [\![E]\!]$. Clearly $\mathbf{V} \subseteq \mathbf{W}$. Let $S \in \mathbf{W}$. Since S is finite, there exists a finite alphabet A and a surjective morphism $\varphi : A^+ \to S$ which can be extended to a uniformly continuous morphism from $\widehat{A^+}$ onto S. Let $\pi_{\mathbf{V}} : \widehat{A^+} \to F_{\mathbf{V}}(A)$ be the natural morphism and let $u, v \in \widehat{A^+}$. By Proposition 3.5, if $\pi_{\mathbf{V}}(u) = \pi_{\mathbf{V}}(v)$, then $u = v$ is an identity of $\mathbf{V}$ and thus, is satisfied by S. In particular, $\pi_{\mathbf{V}}(u) = \pi_{\mathbf{V}}(v)$ implies $\varphi(u) = \varphi(v)$ and by Proposition II.2.11, there exists a morphism $\hat{\pi} : \hat{F}_{\mathbf{V}}(A) \to S$ such that $\varphi = \hat{\pi} \circ \pi_{\mathbf{V}}$.

We claim that $\hat{\pi}$ is uniformly continuous. Since $\hat{F}_{\mathbf{V}}(A)$ is compact by Proposition 3.4, it suffices to verify that $\hat{\pi}$ is continuous. Let F be a subset of the discrete semigroup S. We first observe that $\hat{\pi}^{-1}(F) = \pi_{\mathbf{V}}(\varphi^{-1}(F))$. Since φ is continuous, $\varphi^{-1}(F)$ is closed. Now, $\widehat{A^+}$ is compact, $\pi_{\mathbf{V}}$ is continuous, and $\hat{F}_{\mathbf{V}}(A)$ is Hausdorff. It follows that $\pi_{\mathbf{V}}(\varphi^{-1}(F))$ is closed, proving the claim. It now follows from Proposition 3.4 that S is in $\mathbf{V}$. Thus $\mathbf{V} = \mathbf{W}$. □

For instance, the semigroup $U_1 = \{0, 1\}$, equipped with the usual multiplication of integers, generates the variety of finite idempotent and commutative semigroups, defined by the identities $xy = yx$ and $x = x^2$.

Theorem 3.7 is thus formally similar to Theorem 1.5. The difference lies in the definition of the identities. In Theorem 1.5, an identity is a pair (u, v) of words of A^+ while in Theorem 3.7, an identity is a pair (u, v) of elements of $\widehat{A^+}$.

Each element of $\widehat{A^+}$ is the limit of a Cauchy sequence of (A^+, d). The most important example of a converging sequence is given by the next lemma.

Lemma 3.8 *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 semigroup of size $\leqslant n$. Let indeed $\varphi : \widehat{A^+} \to S$ be a morphism such that $\mathrm{Card}(S) \leqslant n$, and put $s = \varphi(x)$. By Proposition II.5.1, 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 S 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. □

The limit of the sequence $(x^{n!})_{n\geqslant 0}$ is denoted x^ω. In practice, it suffices to remember that its image under any morphism $\varphi : \widehat{A^+} \to S$ onto a finite semigroup is the unique idempotent of the subsemigroup of S generated by $\varphi(x)$. In particular, one can write $\varphi(x^\omega) = \varphi(x)^\omega$ if the integer π on the right hand side is interpreted as the exponent of S.

4 Examples of varieties of finite semigroups

Example 4.1 The variety $[\![x^\omega y x^\omega = x^\omega]\!]$ is the class of semigroups S such that, for each $s \in S$ and for each idempotent $e \in S$, $ese = e$.

Example 4.2 A semigroup belongs to the variety $\mathbf{G_S}$ if and only if it satisfies the identity $x^\omega y = yx^\omega = y$.

A semigroup is *aperiodic* if and only if it satisfies the identity $x^\omega = x^\omega x$, which can also be written, by abuse of notation, $x^\omega = x^{\omega+1}$. Other characterizations of aperiodic semigroups were given in Proposition V.3.9.

A semigroup S is *nilpotent* if and only if, for each idempotent e of S and for each $s \in S$, $es = e = se$. Thus the nilpotent semigroups form a variety, denoted by **Nil**, and defined by the identities $x^\omega y = x^\omega = yx^\omega$. In particular, a nilpotent semigroup is aperiodic. The next proposition summarizes some characteristic properties of the nilpotent semigroups.

Proposition 4.1 *Let S be a nonempty semigroup. The following conditions are equivalent:*

(1) *S is nilpotent,*
(2) *S has a zero which is the unique idempotent of S,*
(3) *there exists an integer $n > 0$ such that S satisfies the identity $x_1 \cdots x_n = y_1 \cdots y_n$.*

Proof. (1) implies (2). Since S is nonempty, Proposition II.5.1 shows that S contains an idempotent e. If S is nilpotent, we have $es = e = se$ for each $s \in S$, showing that e is a zero. Moreover, for each idempotent f of S, we also have

$fs = f = sf$, whence in particular $fe = f$. Since e is a zero, it follows that $f = e$.

(2) implies (3). Let $n = \mathrm{Card}(S)$ and let $s_1, \ldots, s_n$ be a finite sequence of elements of S. By Proposition II.5.4, there exists an idempotent $e \in S$ such that $s_1 \cdots s_i e s_{i+1} \cdots s_n = s_1 \cdots s_n$. Since the unique idempotent of S is a zero, we have $s_1 \cdots s_n = 0$. Therefore S satisfies the identity $x_1 \cdots x_n = y_1 \cdots y_n$.

(3) implies (1). Let $s \in S$ and $e \in E(S)$. Setting $x_1 = s$ and $x_2 = \ldots = x_n = y_1 = \ldots = y_n = e$, we find $s = e$ if $n = 1$ and $se = e$ if $n \geqslant 2$ and hence $se = e$ in all cases. One would show in a similar way that $es = e$ and hence S is nilpotent. □

A finite semigroup is *aperiodic* if there is an integer $n > 0$ such that, for each $x \in S$, $x^n = x^{n+1}$. Since we assume finiteness, quantifiers can be inverted in the definition: S is aperiodic if there is, for each $x \in S$, there is an integer $n > 0$ such that $x^n = x^{n+1}$. Other characterizations of aperiodic semigroups were given in Proposition V.3.9.

We denote by **A** the variety of finite aperiodic semigroups and by by **J** (resp. **R**, **L**), the variety of finite $\mathcal{J}$-trivial (resp. $\mathcal{R}$-trivial, $\mathcal{L}$-trivial) semigroups. The identities defining these varieties are given in the next proposition.

Proposition 4.2 *The following equalities hold*

$$\begin{aligned}
\mathbf{R} &= [\![(xy)^\omega x = (xy)^\omega]\!] \\
\mathbf{L} &= [\![y(xy)^\omega = (xy)^\omega]\!] \\
\mathbf{J} &= [\![y(xy)^\omega = (xy)^\omega = (xy)^\omega x]\!] = [\![x^\omega x = x^\omega, (xy)^\omega = (yx)^\omega]\!] \\
\mathbf{A} &= [\![x^\omega = x^{\omega+1}]\!]
\end{aligned}$$

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 π is interpreted as the exponent of M, we observe that $(xy)^\omega x \;\mathcal{R}\; (xy)^\omega$ since $((xy)^\omega x)(y(xy)^{\pi-1} = (xy)^{2\pi} = (xy)^\omega$. 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 **L**.

(3) Since $\mathbf{J} = \mathbf{R} \cap \mathbf{L}$, it follows from (1) and (2) that **J** is defined by the identities $y(xy)^\omega = (xy)^\omega = (xy)^\omega x$. Taking $y = 1$, we obtain $x^\pi = x^\pi x$ and also $(xy)^\omega = y(xy)^\omega = (yx)^\omega y = (yx)^\omega$. Conversely, suppose that a monoid satisfies the identities $x^\omega x = x^\omega$ and $(xy)^\omega = (yx)^\omega$. Then we have $(xy)^\omega = (yx)^\pi = (yx)^{\omega+1} = y(xy)^\omega x$, 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 $\mathbf{J} \subset \mathbf{L} \subset \mathbf{A}$.

Chapter VII

Star-free languages

The characterization of star-free languages, obtained by Schützenberger in 1965, is, after to Kleene's theorem, the most important result of the theory of finite automata.

1 Star-free languages

Let A be a finite alphabet. The set of *star-free* subsets of A^* is the smallest set $\mathcal{R}$ of subsets 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 subsets follows the same definition scheme as the one of rational subsets, with the difference that the star operation is replaced by the complement. Since the rational subsets are closed under complement, every star-free subset is rational, but we shall see later on that the converse is not true. It follows also immediately from the definition that every finite set is star-free.

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

$$B^* = A^* \setminus \sum_{a \in A \setminus B} A^* a A^* = \Big(\sum_{a \in A \setminus B} 0^c a 0^c \Big)^c$$

(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 Schützenberger's theorem

Recall that a finite monoid M is aperiodic if there exists an integer n such that, for all $x \in M$, $x^n = x^{n+1}$.

Proposition 2.1 *Aperiodic monoids form a variety of finite monoids.*

We also prove a useful property of aperiodic monoids.

Proposition 2.2 *For 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, this identity implies that

$$x^n = x^{2n} \leqslant x^{2n-1} \leqslant x^{2n-2} \leqslant \ldots \leqslant x^{n+1} \leqslant x^n$$

whence $x^n = x^{n+1}$ for all $x \in M$. Thus M is aperiodic. □

We are now ready to state Schützenberger's theorem.

Theorem 2.3 (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 recognizable subsets of A^* and let $L = L_0L_1$. Let M_0, M_1 and M be the ordered syntactic monoids of L_0, L_1 and L.

Lemma 2.4 *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 be a multiple of m such that $n \geqslant n_0 + n_1 + 1$. In particular, $x^n = x^{2n}$ for all $x \in M$. 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 $ux^{n+1}y \in L$. One can of course suppose that $x \neq 1$. If $ux^ny \in L$, there exists a factorization $ux^ny = x_0x_1$ with $x_0 \in L_0$ and $x_1 \in L_1$. Two cases are possible. Either $x_0 = ux^{n_0}r$ with $rx_1 = x^{n-n_0}y$, or $x_1 = sx^{n_1}y$ 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 $xx^{n+1}v \in L$. □

Let us fix an alphabet A and let $\mathcal{A}(A^*)$ be the set of languages of A^* whose syntactic monoide is aperiodic. An elementary computation shows that the

[1] an improved version of this result is given in Theorem X.4.1

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. Further, by Proposition IV.2.3, 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.2.4 and hence under Boolean operations. Lemma 2.4 shows that $\mathcal{A}(A^*)$ is also closed under product. Consequently, $\mathcal{A}(A^*)$ contains the star-free sets.

To establish the converse, we need two elementary properties of aperiodic monoids. The first property is a simple reformulation of Theorem V.1.8 (5) in the case of aperiodic monoids.

Lemma 2.5 (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 $p^n qr^n = q$. Since M is aperiodic, we have $p^n = p^{n+1}$ and hence $pq = pp^n qr^n = p^n qr^n = q$ 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 2.6 *Let M be an aperiodic monoid and let $m \in M$. Then $\{m\} = (mM \cap Mm) \setminus J_m$, with $J_m = \{s \in M \mid m \notin MsM\}$.*

Proof. It is clear that $m \in (mM \cap Mm) \setminus J_m$. Conversely, if $s \in (mM \cap Mm) \setminus J_m$, 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.1.8 that $m \mathcal{H} s$ and hence $m = s$ since M is aperiodic. □

We now need proving that if $\varphi : A^* \to M$ is a morphism from A^* into 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 one to assume that $P = \{m\}$. We shall show that $\varphi^{-1}(m)$ is star-free by induction on the integer $r(m) = \mathrm{Card}(M \setminus MmM)$. The intial step is treated in the next lemma.

Lemma 2.7 *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 $umv = 1$. 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 occurring in 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

$$\varphi^{-1}(m) = (UA^* \cap A^*V) \setminus (A^*CA^* \cup A^*WA^*) \tag{2.1}$$

where

$$U = \sum_{(n,a)\in E} \varphi^{-1}(n)a \qquad V = \sum_{(a,n)\in F} a\varphi^{-1}(n)$$
$$C = \{a \in A \mid m \notin M\varphi(a)M\} \quad W = \sum_{(a,n,b)\in G} a\varphi^{-1}(n)b$$

with

$$E = \{(n,a) \in M \times A \mid n\varphi(a)\ \mathcal{R}\ m \text{ but } n \notin mM\}$$
$$F = \{(a,n) \in A \times M \mid \varphi(a)n\ \mathcal{L}\ m \text{ but } n \notin Mm\}$$
$$G = \{(a,n,b) \in (A \times M \times A \mid m \in (M\varphi(a)nM \cap Mn\varphi(b)M) \setminus M\varphi(a)n\varphi(b)M\}$$

Denote by L the right hand side of (2.1). 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, $(n,a) \in E$ and $u \in \varphi^{-1}(n)aA^*$. Therefore $u \in UA^*$ and a symmetric argument would show 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 with 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 with the definition of G. Therefore $u \in L$.

Conversely, let $u \in L$ and let $s = \varphi(u)$. Since $u \in UA^* \cap A^*V$, we have $s \in mM \cap Mm$. By Lemma 2.6, 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 $n \in G$ and $f \in W$, which is impossible again.

Formula (2.1) 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.1.8 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 would work for V. There remains to treat the case 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.1.9, $n\varphi(b)\ \mathcal{L}\ \varphi(a)n\varphi(b)$, and hence $m\ \mathcal{J}\ \varphi(a)n\varphi(b)$, a contradiction with the definition of G. Consequently $r(n) < r(m)$ and W is star-free by the induction hypothesis. □

Example 2.1 Let $A = \{a,b\}$ and let $L = (ab)^*$. The minimal automaton of L is represented in Figure 2.1.

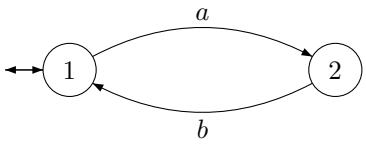


Figure 2.1. The minimal automaton of $(ab)^*$.

The syntactic monoid M of L is the monoid consisting of the six matrices

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad a = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \qquad b = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$$
$$aa = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \qquad ab = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \qquad ba = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

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 2.2:

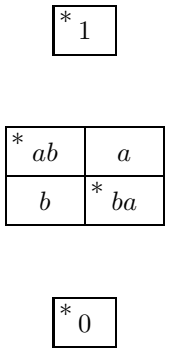


Figure 2.2. The $\mathcal{D}$-class structure of M.

This monoid is aperiodic, since $x^2 = x^3$ for each $x \in M$, and hence L is star-free.

Example 2.2 Let $A = \{a, b\}$ and let $L' = (aa)^*$. The minimal automaton of L' is represented in Figure 2.3:

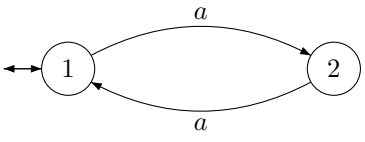


Figure 2.3. The minimal automaton of $(aa)^*$.

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 2.4:

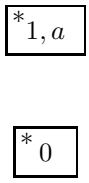


Figure 2.4. The $\mathcal{D}$-class structure of M'.

This monoid is not aperiodic, since, for each $n > 0$, $a^n \neq a^{n+1}$ and hence L' is not star-free.

Chapter VIII

Piecewise testable languages

Simon's theorem shows that the languages recognized by $\mathcal{J}$-trivial monoids are exactly the shuffle ideals. This result has far reaching consequences, both in semigroup theory and in automata theory. We shall present two of the seven published proofs of Simon's theorem: the first one, due to Imre Simon, is based on a careful analysis of the subword ordering and has a strong combinatorial flavour. The second one, due to Straubing and Thérien, is more algebraic in nature.

As a preliminary step, we shall explore the properties of the subword ordering and give an algebraic characterization 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:

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 a subword-free sequence of words,

(3) u_2 is a shortest word such that u_0, u_1, u_2 is the beginning 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 $i_0 < i_1 < \ldots$ Now the sequence $u_0, u_1, \ldots, u_{i_0-1}, v_{i_0}, v_{i_1}, \ldots$ 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 subword 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 less $\leqslant n$. The word w can be factorized 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 denote by $c(u)$ the *content* of u, that is, the set of letters of A occurring in u. For instance, $c(babaa) = \{a, b\}$.

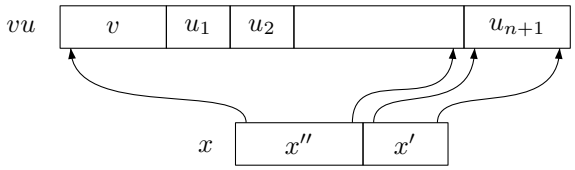
Proposition 1.4 *Let $u, v \in A^*$ and let $n > 0$. Then $u \sim_n vu$ if and only if there exist $u_1, \ldots, u_n \in A^*$ such that $u = u_1 \cdots u_n$ and $c(v) \subseteq c(u_1) \subseteq \ldots \subseteq c(u_n)$.*

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. Further, 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. □

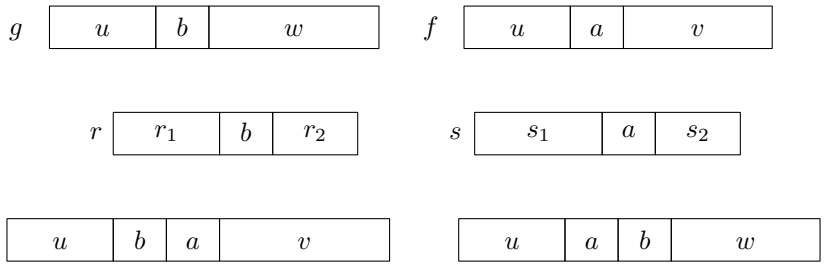
Corollary 1.5 *For every $u, v \in A^*$, one has $(uv)^n u \sim_n (uv)^n \sim_n v(uv)^n$.*

Proof. The formula $(uv)^n \sim_n v(uv)^n$ follows from Proposition 1.3. 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 of length $\leqslant n$ which is a subword of $uabw$ but not of uaw.

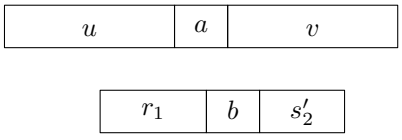


Necessarily $r = r_1br_2$ where r_1 is a subword of u and r_2 is a subword of av, and $s = s_1as_2$ where s_1 is a subword of u and s_2 is a subword of bw. It follows that r_1b is not a subword of u (for otherwise $r = r_1br_2$ would be a subword of $uav = f$ and therefore of g). Likewise s_1a is not 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

$$|r_1bs_2'| + |s_1ar_2'| \leqslant |r_1as_2| + |s_1br_2| \leqslant |r| + |s| \leqslant 2n$$

and therefore one of the words r_1bs_2' or s_1ar_2' is of length $\leqslant n$. Suppose for example that this is r_1bs_2'. Then r_1bs_2' is a subword of $ubw = g$ and therefore also of $f = uav$. However, r_1b is not a subword of u. Thus bs_2' is a subword of v, and a fortiori s_2 is a subword of v.



Thus $s = s_1as_2$ is a subword of $uab = f$, 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 $f = a^3b^3a^3b^3$ and $g = a^2b^4a^4b^2$. 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

$$f = (aa)a(b^3a^3b^3) \sim_4 (aa)b(b^3a^4b^2) = g$$

whence

$$(aa)a(b^3a^3b^3) \sim_4 (aa)ab(b^3a^4b^2) \quad \text{or} \quad (aa)b(b^3a^4b^2) \sim_4 (aa)ba(b^3a^3b^3)$$

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 $a_1, \dots, a_k$. 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 characterization of shuffle ideals:

Theorem 2.1 *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 L, 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 $u \in F$. □

One can give a constructive proof which does not rely on Higman's theorem.

Proposition 2.2 *Let L be recognizable language such that $L \text{ ш } A^* = L$. Then one can effectively find a finite set of words F such that $L = F \text{ ш } A^*$.*

Proof. Let n be the number of states of the minimal automaton $\mathcal{A}$ of L. Set

$$F = \{u \in L \mid |u| \leqslant n\} \text{ and } K = F \text{ ш } A^*$$

We claim that $L = K$. Since $F \subseteq L$, one has $F \text{ ш } A^* \subseteq L \text{ ш } A^* = L$ and hence $K \subseteq L$. If the inclusion is strict, consider a word u of minimal length in $L \setminus K$. Necessarily, $|u| > n$, for otherwise $u \in F$. Let $u = a_1 \cdots a_r$ and let $q_1 \xrightarrow{a_1} q_2 \xrightarrow{a_2} q_3 \cdots q_{r-1} \xrightarrow{a_{r-1}} q_r$ be a successful path of 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 $K \text{ ш } A^*$. Now, since

$$K \text{ ш } A^* = (F \text{ ш } A^*) \text{ ш } A^* = F \text{ ш } (A^* \text{ ш } A^*) = F \text{ ш } A^* = K$$

one has $u \in K$, a contradiction. This proves the claim and the proposition. □

Corollary 2.3 *Every shuffle ideal is a recognizable language.*

We now come to the algebraic characterization of shuffle ideals.

Theorem 2.4 *A language is a shuffle ideal if and only if its ordered syntactic monoid satisfies the identity $x \leqslant 1$.*

Proof. Let L be a language and let $\eta : A^* \to (M, \leqslant)$ be its ordered syntactic morphism. Suppose that L is a shuffle ideal. If $uv \in L$, then $uxv \in L$ for each $x \in A^*$. Therefore $x \leqslant_L 1$ and thus M satisfies the identity $x \leqslant 1$.

Conversely, if M satisfies the identity $x \leqslant 1$, then, for every $x \in A^*$, $x \leqslant_L 1$, 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 terminilogy 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 3.1 *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 $L = A^*a_1A^* \cdots a_nA^*$ 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 reflexion should suffice to verify the following formula:

$$\{v \in A^* \mid v \sim_n u\} = \Big(\bigcap_{a_1 \cdots a_k \in E} A^*a_1A^* \cdots a_kA^* \Big) \setminus \Big(\bigcup_{a_1 \cdots a_k \in F} A^*a_1A^* \cdots a_kA^* \Big)$$

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 characterization of piecewise testable languages is the main result of this chapter. It relies on two results of semigroup theory of independent interest.

Proposition 3.2 *Any finite ordered monoid satisfying the identity $x \leqslant 1$ 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 $r \leqslant 1$ and $s \leqslant 1$, it follows that $x = rys \leqslant y$ and similarly, $y \leqslant x$. Thus $x = y$. □

Theorem 3.3 (Simon) *Let M be a finite $\mathcal{J}$-trivial monoid and let n is 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.2.11, 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 $v = v_1v_2 \cdots v_n$ such that $\{a\} \subseteq c(v_1) \subseteq \ldots \subseteq c(v_n)$. Consider the $\leqslant_{\mathcal{J}}$-chain of length $n + 1$

$$\varphi(v_1 \cdots v_n) \leqslant_{\mathcal{J}} \varphi(v_2 \cdots v_n) \leqslant_{\mathcal{J}} \cdots \leqslant_{\mathcal{J}} \varphi(v_1) \leqslant_{\mathcal{J}} 1$$

By the choice of n, this chain is not strict and there exist two indices $i < j$ such that $\varphi(v_i \cdots v_n) \mathcal{J} \varphi(v_j \cdots v_n)$. Since M is $\mathcal{J}$-trivial, one has $\varphi(v_i \cdots v_n) = \varphi(v_j \cdots v_n) = s$. Let $b \in c(v_i)$. Then $v = v_i'bv_i''$ for some $v_i', v_i'' \in A^*$ and thus $s = \varphi(v_i \cdots v_n) \mathcal{J} \varphi(bv_i''v_{i+1} \cdots v_j \cdots v_n) \mathcal{J} \varphi(v_i''v_{i+1} \cdots v_j \cdots v_n) \mathcal{J} \varphi(v_j \cdots v_n) = s$. Consequently, $\varphi(b)s = s$ for each $b \in c(v_i)$ and therefore $\varphi(v) = \varphi(v_1 \cdots v_n) = s = \varphi(a)s = \varphi(av)$. It follows that $\varphi(f) = \varphi(uav) = \varphi(uv) = \varphi(g)$, which concludes the proof. □

Theorem 3.3 has a very important consequence.

Corollary 3.4 *Every finite monoid is the quotient of a finite ordered monoid satisfying the identity* $x \leqslant 1$.

Proof. Indeed, the subword ordering induces a stable partial order on $A^*/\sim_{2n-1}$. Furthermore, since the empty word is a subword of every word, the identity $1 \leqslant x$ holds in this ordered monoid. □

We now return to the announced characterization of piecewise testable languages.

Theorem 3.5 (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 2.4, the ordered syntactic monoid of L satisfies the identity $x \leqslant 1$. By Proposition 3.2, this monoid is $\mathcal{J}$-trivial. Now if L is piecewise testable, it is by Proposition 3.1 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 3.3, 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 $\mathfrak{T}_n$ consisting of all order preserving and extensive functions from $\{1, \ldots, n\}$ into 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 one entries on the diagonal. For example

$$\mathcal{R}_2 = \{\left(\begin{smallmatrix}1&0\\0&1\end{smallmatrix}\right), \left(\begin{smallmatrix}1&1\\0&1\end{smallmatrix}\right), \left(\begin{smallmatrix}1&0\\1&1\end{smallmatrix}\right), \left(\begin{smallmatrix}1&1\\1&1\end{smallmatrix}\right)\}$$

Finally, $\mathcal{U}_n$ is the submonoid of $\mathcal{C}_n$ consisting of the upper triangular matrices of $\mathcal{C}_n$. The matrices of $\mathcal{U}_n$ are called *unitriangular*. For example,

$$\mathcal{U}_3 = \left\{ \left(\begin{smallmatrix}1&\varepsilon_1&\varepsilon_2\\0&1&\varepsilon_3\\0&0&1\end{smallmatrix}\right) \mid \varepsilon_1, \varepsilon_2, \varepsilon_3 \in \{0, 1\} \right\}$$

Proposition 4.1 *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 a 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 3.2. □

The next proposition is another elementary property of the monoids $\mathcal{C}_n$, $\mathcal{R}_n$ and $\mathcal{U}_n$.

Proposition 4.2 *For each $n, m > 0$, the monoids $\mathcal{C}_n \times \mathcal{C}_m$, $\mathcal{R}_n \times \mathcal{R}_m$ and $\mathcal{U}_n \times \mathcal{U}_m$ is isomorphic to a submonoid $\mathcal{C}_{n+m}$, $\mathcal{R}_{n+m}$ and $\mathcal{U}_{n+m}$ the monoids, respectively.*

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 vareity of $\mathcal{J}$-trivial monoids.

Theorem 4.3 *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 4.1, the monoids $\mathcal{C}_n$, $\mathcal{R}_n$ and $\mathcal{U}_n$ are $\mathcal{J}$-trivial. Therefore one 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. TO DO. □

Theorem 4.4 *Every finite $\mathcal{J}$-trivial monoid is a quotient of a finite ordered monoid satisfying the identity $x \leqslant 1$.*

Proof. TO DO. □

Chapter IX

The variety theorem

A *class* of recognizable subsets is a correspondence that associates with each finite alphabet A, a set $\mathcal{C}(A^*)$ of recognizable languages of A^*. 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".

One can associate with each variety of monoids $\mathbf{V}$ the class $\mathcal{V}$ of languages recognized by a monoid of $\mathbf{V}$. Such a class $\mathcal{V}$ is called a variety of languages and admits an abstract characterization: it is a class of recognizable languages closed under Boolean operations, inverse morphisms between free monoids, and right and left quotients. Eilenberg's variety theorem asserts that the correspondence $\mathbf{V} \to \mathcal{V}$ between varieties of finite monoids and varieties of languages is bijective. For instance, the variety of rational languages corresponds to the variety of finite monoids, the variety of star-free languages corresponds to the variety of aperiodic monoids, and that of piecewise testable languages corresponds to the variety of $\mathcal{J}$-trivial monoids.

1 Varieties of languages

Let E be a set. A class of subsets of E forms a *positive Boolean algebra* if it is closed under finite union and finite intersection. It contains in particular the empty subset (obtained as the union of the empty family) and the full subset E (obtained as the intersection of the empty family). A *Boolean algebra* is a positive Boolean algebra which is also closed under complementation. The smallest positive Boolean algebra containing a set $\mathcal{E}$ of subsets of E is called the positive Boolean algebra *generated* by $\mathcal{E}$. Its elements are obtained from the elements of $\mathcal{E}$ by using the operations of finite union and finite intersection: they are the *positive Boolean combinations* of elements of $\mathcal{E}$. The notion of generated Boolean algebra can be defined in the same way by adding the complement to the basic operations.

Recall that if X is a subset of A^* and if $u \in A^*$, the *left* (resp. *right*) *quotient* of X by u is the set

$$u^{-1}X = \{v \in A^* \mid uv \in X\} \qquad (\text{resp. } Xu^{-1} = \{v \in A^* \mid vu \in X\})$$

A *positive variety* is a class of recognizable subsets such that

(1) for each alphabet A, $\mathcal{V}(A^*)$ is a positive Boolean algebra,

(2) for each morphism of semigroups $\varphi : A^* \to B^*$, $X \in \mathcal{V}(B^*)$ 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* is a positive variety closed under complementation. This amounts replacing (1) by (1') in the previous definition

(1′) for each alphabet A, $\mathcal{V}(A^*)$ is a Boolean algebra,

We are now ready to state the variety theorem.

Theorem 1.1 *The correspondence $\mathbf{V} \to \mathcal{V}$ defines a bijection between the varieties of ordered semigroups and the +-positive varieties on one hand and between the varieties of semigroups and the +-varieties on the other hand.*

Boldface letters are usually used to denote varieties of finite semigroups and the corresponding +-varieties are denoted with cursive letters.

One can obtain a parallel theory by considering the recognizable subsets of the free semigroup A^+. It suffices to define successively the notions of subset recognized by an (ordered) semigroup, of (ordered) syntactic semigroup, of variety of (ordered) semigroups. The main difference with the corresponding notions for monoids concerns the identities. Consider for instance the identity $x^\omega y x^\omega = x^\omega$. This identity is actually equivalent, in the monoid case, to the identity $y = 1$. To see this, it suffices to substitute x by 1, to obtain $y = 1$. The only monoid satisfying the identity $x^\omega y x^\omega = x^\omega$ is therefore the trivial monoid.

Continuing the parallelism, one next defines positive +-varieties as follows: a *+-positive variety* is a class of recognizable subsets such that

(1) for each alphabet A, $\mathcal{V}(A^+)$ is a positive Boolean algebra,

(2) for each morphism of semigroups $\varphi : A^+ \to B^+$, $X \in \mathcal{V}(B^+)$ 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 $Xu^{-1} \in \mathcal{V}(A^+)$.

Similarly, a *+-variety* is a +-variety closed under complementation. Then we have a semigroup version of Theorem 1.1.

Theorem 1.2 *The correspondence $\mathbf{V} \to \mathcal{V}$ defines a bijection between the varieties of ordered semigroups and the positive +-varieties on the one hand and between the varieties of semigroups and the +-varieties on the other hand.*

The variety theorem allows one to associate a variety (resp. +-variety) with each variety of monoids (resp. semigroups). The most interesting case are of course those for which a combinatorial description of the corresponding variety of languages is known. We shall present several such examples in the next section.

2 Some examples of varieties.

The description of the +-variety associated with the trivial variety of semigroups **1** is immediate.

Proposition 2.1 *For each alphabet A, $\mathcal{I}(A^*)$ consists of the empty and the full subsets.*

We pursue with the varieties **Nil**, $\mathbf{Nil}^+$ and $\mathbf{Nil}^-$. Recall that a subset F of a set E is *cofinite* if the complement of F in E is finite.

Proposition 2.2 *For each alphabet A,*

(1) *$\mathcal{N}il^+(A^+)$ consists of the empty subset and of the cofinite subsets of A^+,*

(2) *$\mathcal{N}il^-(A^+)$ consists of A^+ and of the finite subsets of A^+,*

(3) *$\mathcal{N}il(A^+)$ is the set of finite or cofinite subsets of A^+.*

Proof. (1). Denote by $\varphi : A^+ \to S$ the syntactic morphism of L. If L is empty, S is trivial, and hence in $\mathbf{Nil}^+$. 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 thus have the same image e under φ. By Proposition VI.4.1, S is thus nilpotent. There remains to prove that $e \leqslant s$ for every $s \in S$. Let $v \in \varphi^{-1}(s)$. Then the formal implication

$$(xvy \in L \Rightarrow xuy \in L)$$

shows that $u \leqslant_L v$, whence $e \leqslant s$ in S. Therefore $S \in \mathbf{Nil}^+$.

Conversely, let $(S, \leqslant) \in \mathbf{Nil}^+$, I be an order ideal of S and let $\varphi : A^+ \to S$ be a morphism of semigroups. If I is empty, $\varphi^{-1}(I)$ is empty also. Otherwise, I contains necessarily 0, since 0 is minimal for $\leqslant$. Let u be a word of length greater than or equal to $\mathrm{Card}(S)$. By Proposition VI.4.1, $\varphi(u) = 0$ 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 $S \setminus P$ and the argument above shows that $\varphi^{-1}(P)$ is either finite or cofinite. □

If a variety is generated by a single [ordered] monoid, the corresponding [positive] variety of languages is easy to describe.

Proposition 2.3 *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 every alphabet A, $\mathcal{V}(A^*)$ is the positive Boolean algebra generated by the sets of the form $\varphi^{-1}(\downarrow m)$, where $\varphi : A^* \to M$ is an arbitrary morphism and $m \in M$.*

Proof. It is clear that $\varphi^{-1}(\downarrow m) \in \mathcal{V}(A^*)$ and thus $\mathcal{V}(A^*)$ also contains the positive Boolean algebra 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 order ideal I of M such that $L = \varphi^{-1}(I)$. Since $\varphi^{-1}(I) = \bigcup_{m \in P} \varphi^{-1}(\downarrow m)$, it is sufficient to establish the result when $L = \varphi^{-1}(\downarrow m)$ 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

$$\varphi^{-1}(\downarrow m) = \bigcap_{1 \leqslant i \leqslant n} (\pi_i \circ \varphi)^{-1}(\downarrow m_i)$$

Since $m_i \in M$ and $\pi_i \circ \varphi$ is a morphism from A^* into M, the result follows. □

There is of course a similar result for varieties of monoids, the proof of which is similar.

Proposition 2.4 *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* $m \in M$.

It follows in particular that if a variety $\mathbf{V}$ is generated by a single [ordered] monoid, then, for each alphabet A, the set $\mathcal{V}(A^*)$ is finite. For this reason, the variety $\mathbf{V}$ is called *locally finite*.

Let $\mathbf{J_1}$ be the variety $\mathbf{J_1}$ of idempotent and commutative monoids, defined by the identities $xy = yx$ and $x^2 = x$. By Proposition XII.3.1, $\mathbf{J_1}$ is generated by its cyclic monoids. But there is only one nontrivial cyclic monoid in $\mathbf{J_1}$, and this is the monoid $U_1 = \{0, 1\}$ considered in Section II.1.5. The corresponding variety is described as follows.

Proposition 2.5 *For each alphabet* A, $\mathcal{J}_1(A^*)$ *is the Boolean algebra generated by the subsets of the form* A^*aA^* *where* a *is a letter. Equivalently,* $\mathcal{J}_1(A^*)$ *is the Boolean algebra generated by the subsets 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

$$B^* = A^* \setminus \bigcup_{a \in A \setminus B} A^*aA^* \quad \text{and} \quad A^*aA^* = A^* \setminus (A \setminus \{a\})^*$$

Since $\mathbf{J_1}$ is generated by U_1, one can use Proposition 2.4 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 $\varphi^{-1}(0) = A^* \setminus B^*$, which establishes the proposition. □

This result can be divided into two parts. Denote by $\mathbf{J_1^+}$ (resp. $\mathbf{J_1^-}$) the positive variety of idempotent and commutative monoids satisfying the identity $x \leqslant 1$. If B is a subset of A, denote by $F(B)$ the set of words of A^+ containing at least one occurrence of each letter of B. Thus

$$F(B) = \bigcap_{a \in B} A^*aA^*$$

The next proposition is thus a variant of Proposition 2.5 and its proof is left as an exercise to the reader.

Proposition 2.6 *For each alphabet* A, $\mathcal{J}_1^+(A^*)$ *is the set of finite unions of subsets of the form* $F(B)$ *where* $B \subset A$. *Similarly,* $\mathcal{J}_1^-(A^*)$ *is the set of finite unions of subsets of the form* B^+ *where* $B \subset A$.

Another interesting example of locally finite variety is the variety $\mathbf{R_1}$ of idempotent and $\mathcal{R}$-trivial monoids.

Proposition 2.7 *Let L be a recognizable 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 $B \subset A$,*

Proof. (1) implies (2) since $\tilde{U}_2 \in \mathbf{R_1}$.
(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 sequential function, realized by the sequential transducer $\mathcal{T} = (\mathcal{P}(A), A, \emptyset, ., *)$, 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

$$L_{(a_1,\ldots,a_n)} = A_n^* \cap \bigcap_{1\leqslant i\leqslant n} A_{i-1}^*a_iA_i^* \text{ where } A_i = \{a_1,\ldots,a_i\} \text{ and } A_0 = \emptyset$$

Condition (5) is now a consequence of the following equalities:

$$A_i^* = A^* \setminus \bigcup_{a\notin A_i} A^*aA^* \qquad A_{i-1}^*a_iA_i^* = A_{i-1}^*a_iA^* \cap A_i^*$$

(5) implies (1). By the variety theorem (Theorem 1.1), it is sufficient to show that, for $B \subset A$ and $a \in A$, B^*aA^* is recognized 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

$$\varphi(a) = a_1$$
$$\varphi(b) = \begin{cases} 1 & \text{if } b \in B \setminus \{a\} \\ a_2 & \text{for } b \in A \setminus (B \cup \{a\}) \end{cases}$$

Then $\varphi^{-1}(a_1) = B^*aA^*$, which concludes the proof. □

We shall conclude this section with a description of the $*$-variety corresponding to the variety **Acom** of aperiodic and commutative monoids.

If a is a letter of an alphabet A, let us denote by $L(a, k)$ 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 2.8 *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 recognized 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 XII.3.1, **Acom** is generated by its cyclic monoids, and Proposition 2.4 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

$$\varphi^{-1}(x^n) = A^* \setminus \bigcup_{0 \leqslant k < n} \varphi^{-1}(x^k)$$

which concludes the proof. □

The original version of the variety theorem dealt only with semigroup varieties. Its extension to ordered semigroups is due to Pin [24].

Chapter X

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 proof of the next result is immediate.

Proposition 1.1 *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 1.2 *Let $\tau : S \to T$ be a relational morphism. Then the graph R of τ is a subsemigroup of $S \times T$ and 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 factorization 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 factorization $\tau = \beta \circ \alpha^{-1}$, pictured in Figure 1.1 is called the *canonical factorization* of τ.

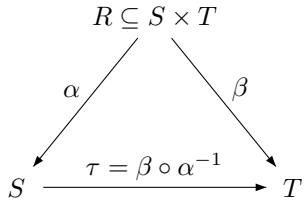


Figure 1.1. The canonical factorization of a relational morphism.

We shall see that in most cases the properties of τ are bounded to that of β (see in particular Propositions 2.1 and 3.3).

The next result extends Proposition to II.2.1 to relational morphisms. We remind the reader that if τ is a relation from S into T and T' is a subset of T, then $\tau^{-1}(T') = \{s \in S \mid \tau(s) \cap T \neq \emptyset\}$.

Proposition 1.3 *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 $\{1, 2, 3, 4\}$ into 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 $\tau^{-1}(Id)$ 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 2.1 *Let $S \xrightarrow{\alpha^{-1}} R \xrightarrow{\beta} T$ be the canonical factorization of a relational morphism $\tau : S \to T$. Then τ is injective (resp. surjective) if and only if β is injective (resp. 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 (resp. surjective), then $\tau = \beta \circ \alpha^{-1}$ is also injective (resp. 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 2.1 has two interesting consequences.

Corollary 2.2 *A semigroup S divides a semigroup T if and only if there exists an injective relational morphism from S into 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 into T.

Conversely, if τ is an injective relational morphism from S into T and if $S \xrightarrow{\alpha^{-1}} R \xrightarrow{\beta} T$ is the canonical factorization of τ. Proposition 2.1 shows that β is injective. Since α is surjective, S divides T. □

Corollary 2.3 *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 factorization of τ. Then β is injective by Proposition 2.1 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 $e \in \tau(s) \cap \tau(s^2)$ 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 2.3 can be improved as follows.

Proposition 2.4 *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 1.3, $\tau(e)$ is a subsemigroup of T, which, by Corollary II.5.2, 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 2.3. □

3 Relational V-morphisms

All semigroups considered in this section are finite.

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 the following varieties:

(1) $\mathbf{A}$, the variety of aperiodic semigroups,

(2) $\mathbf{LI} = [\![ese = e]\!]$, the variety of locally trivial semigroups,

(3) $\mathbf{LJ}^+ = [\![ese \leqslant e]\!]$, the variety of ordered semigroups S, such that, for all $e \in E(S)$, the ordered submonoid eSe satisfies the identity $x \leqslant 1$.

A relational $\mathbf{A}$-morphism is also called an *aperiodic relational morphism* and a relational $\mathbf{LI}$-morphism is also called a *locally trivial relational morphism*.

The definition of 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 3.1 *Relational $\mathbf{V}$-morphisms are closed under composition.*

Let us mention another elementary result.

Proposition 3.2 *Injective relational morphisms are relational $\mathbf{V}$-morphisms for every variety $\mathbf{V}$.*

Proof. This follows directly from Corollary 2.3. □

Note that the converse to Proposition 3.2 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 **V**-morphism for every variety **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 factorizations.

Proposition 3.3 *Let $S \xrightarrow{\alpha^{-1}} R \xrightarrow{\beta} T$ be the canonical factorization of a relational morphism $\tau : S \to T$. Then τ is a relational* **V***-morphism if and only if β is a* **V***-morphism.*

Proof. First, α^{-1} is an injective relational morphism and thus a relational **V**-morphism by Proposition 3.2. Thus if β is a relational **V**-morphism, then τ is a relational **V**-morphism by Proposition 3.1.

Conversely, suppose that τ is a relational **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 **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 **V**. Thus β is a relational **V**-morphism. □

Relational morphisms can be restricted to subsemigroups.

Proposition 3.4 *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 (resp. a relational* **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 **V**-morphisms.

3.1 Aperiodic relational morphisms

Theorem 3.5 *Let $S \xrightarrow{\alpha^{-1}} R \xrightarrow{\beta} T$ be the canonical factorization 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 3.3. 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 Proposition V.7.8 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 of 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 first theorem to β gives the equivalence of (1')–(4'). □

3.2 Locally trivial relational morphisms

Theorem 3.6 *Let $S \xrightarrow{\alpha^{-1}} R \xrightarrow{\beta} T$ be the canonical factorization 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 3.3. 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 3.4, we may assume that τ is surjective. Further, it follows from Theorem 3.5 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 would show that $f \leqslant_{\mathcal{J}} e$ and hence $e\{$. 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 first theorem to β gives the equivalence of (1')–(2'). □

Proposition 3.7 *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 $\mathcal{D}$-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 $\pi^{-1}(e)$, 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 order ideal generated by an element $x \in S$ is the set $\downarrow x$ of all $y \in E$ such that $y \leqslant x$.

Proposition 3.8 *Let S be an ordered semigroup and let $e \in E(S)$. Then the ordered semigroup $e(\downarrow e)e$ belongs to the variety $[\![ese \leqslant e]\!]$.*

Proof. Let $R = e(\downarrow e)e$. Let $r \in R$ and $f \in E(R)$. Then $f = ege$ with $g \leqslant e$ and $r = ese$ with $s \leqslant e$. It follows $ef = f = fe$ and $frf = fesef = fsf \leqslant fef = f$. Thus $R \in [\![ese \leqslant e]\!]$. □

Proposition 3.9 *Let $\tau : S \to T$ be a relational morphism. The following conditions are equivalent:*

(1) *τ is a relational $[\![ese \leqslant e]\!]$-morphism,*
(2) *for any $e \in E(T)$, $\tau^{-1}(e(\downarrow e)e)$ is an ordered semigroup of $[\![ese \leqslant e]\!]$,*
(3) *for any $e \in E(T)$, $f \in E(\tau^{-1}(e))$ and $s \in \tau^{-1}(e(\downarrow e)e)$, $fsf \leqslant f$.*

Proof. Proposition 3.8 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 $R \in [\![ese \leqslant e]\!]$. 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 $ere \leqslant e$ since $R \in [\![ese \leqslant e]\!]$ and thus $e, ere \in e(\downarrow e)e$. 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 $fsf \leqslant f$ and thus $U \in [\![ese \leqslant e]\!]$. Therefore, τ is a relational $[\![ese \leqslant e]\!]$-morphism. □

4 Three examples of relational morphisms

In this section, we give three examples from the theory of automata and recognizable 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 properties of the flower automata.

4.1 Concatenation product

Let, 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

$$\eta : A^* \to M(L_0) \times M(L_1) \times \cdots \times M(L_n)$$

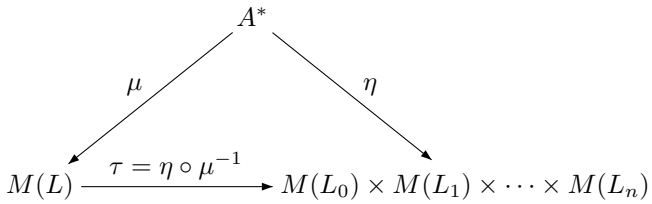
be the morphism defined by

$$\eta(u) = (\eta_0(u), \eta_1(u), \ldots, \eta_n(u))$$

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 [35] and later in [22, 29, 26].



Theorem 4.1 *The relational morphism $\tau : M(L) \to M(L_0) \times M(L_1) \times \cdots \times M(L_n)$ is a relational $[\![ese \leqslant e]\!]$-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 $x^\omega y x^\omega \leqslant x^\omega$, 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 $u, v \in A^*$, $ux^k v \in L$ implies $ux^k y x^k v \in L$. Let $r = 3n+1$. Then $\eta(x^{rk}) = \eta(x^k)$, and since $ux^k v \in L$, $ux^{rk} v \in L$. Consequently, there is a factorization of the form $ux^{rk}v = u_0 a_1 \cdots a_n u_n$, where $u_i \in L_i$ for $0 \leqslant i \leqslant n$. We claim that one of the factors u_h contains x^{2k} as a factor. Otherwise the length of each u_h would be at most $3k|x| - 2$ and the following sequence of inequalities, which follows from the choice of r,

$$rk|x| \leqslant |ux^{rk}v| = |u_0 a_1 \cdots a_n u_n| \leqslant (3k|x| - 2)(n+1) + n < rk|x|$$

would give a contradiction. Therefore, there exist $1 \leqslant h \leqslant n$ and $0 \leqslant j \leqslant r-2$ such that $u_h = u'_h x^{2k} u''_h$ for some $u'_h, u''_h \in A^*$, $ux^{jk} = u_0 a_1 \cdots a_{h-1} u'_h$ and $x^{(r-j-2)k} v = u''_h a_h \cdots a_n u_n$. Now since $\eta(x)$ and $\eta(y)$ belong to R,

$$\eta(x^k)\eta(y)\eta(x^k) \leqslant \eta(x^k)$$

and by projection onto $M(L_h)$, $\eta_h(x^k)\eta_h(y)\eta_h(x^k) \leqslant \eta_h(x^k) = \eta_h(x^{2k})$. In particular, the condition $u'_h x^{2k} u''_h \in L_h$ implies $u'_h x^k y x^k u''_h \in L_h$. It follows $ux^{(j+1)k} y x^{(r-j-1)k} v \in L$, and hence $ux^k y x^k v \in L$, which concludes the proof. □

Theorem 4.1 is often used in the following weaker form.

Corollary 4.2 *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 4.1, τ is a relational $[\![ese \leqslant e]\!]$-morphism. In particular, for each idempotent e, $\tau^{-1}(e)$ is a semigroup satisfying the identity $ese \leqslant e$. In particular, it satisfies the identity $x^\omega x x^\omega \leqslant x^\omega$, that is $x^{\omega+1} \leqslant x^\omega$ and is aperiodic by Proposition VII.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

$$L = L_0 a_1 L_1 \cdots a_n L_n$$

is said to be *unambiguous* if every word of L admits a unique decomposition of the form $u = u_0 a_1 u_1 \cdots a_n u_n$ with $u_0 \in L_0$, ..., $u_n \in L_n$.

Example 4.1 Let $A = \{a, b, c\}$. The marked product $\{a, c\}^* a \{1\} b \{b, c\}^*$ is unambiguous.

Theorem 4.3 *If the product $L_0 a_1 L_1 \cdots a_n L_n$ 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 3.6, 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 4.1 that R satisfies the identity $x^\omega y x^\omega \leqslant x^\omega$ and it just remains to prove the opposite identity $x^\omega \leqslant x^\omega y x^\omega$. Let $x, y \in \eta^{-1}(e)$ and let k be an integer such that $\mu(x^k)$ is idempotent. It suffices to show that $ux^k y x^k v \in L$

implies $ux^kv \in L$. Let $r =$. Then $\eta(x^{rk}) = \eta(x^k)$, and if $ux^kyx^kv \in L$, then $ux^{rk}yx^{rk}v \in L$. Consequently, there is a factorization of the form $ux^{rk}yx^{rk}v = u_0a_1 \cdots a_nu_n$, where $u_i \in L_i$ for $0 \leqslant i \leqslant n$.

First assume that one of the words u_i contains xyx as a factor, that is, $u_i = u_i'xyxu_i''$ with $u_0a_1 \cdots a_iu_i' = ux^{rk-1}$ and $u_i''a_{i+1} \cdots a_nu_n = x^{rk-1}v$. Since $\eta(x) = \eta(y) = e$, one has $\eta_i(xyx) = \eta_i(x) = \eta_i(x^2)$ and hence, $u_i'xyxu_i'' \in L_i$ implies $u_i'x^2u_i'' \in L_i$. Consequently, one has

$$ux^{2rk}v = ux^{rk-1}x^2x^{rk-1}v = u_0a_1 \cdots a_i(u_i'x^2u_i'')a_{i+1} \cdots a_nu_n$$

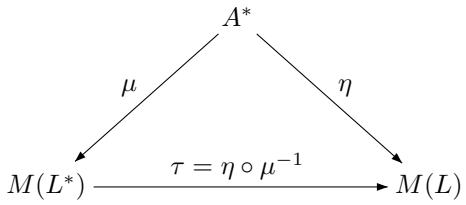
which shows that $ux^{2rk}v$ belongs to L. Since $\eta(x^{2rk}) = \eta(x^k)$, it follows that ux^kv is also in L, as required.

Suppose now that none of the words u_i contains xyx as a factor. □

4.2 Pure languages

A submonoid L^* of A^* is *pure* if for all $u \in A^*$ and $n > 0$, the condition $u^n \in L^*$ implies $u \in L^*$.

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 $M(L)$.



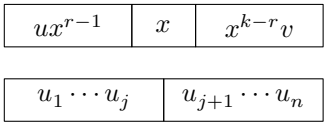
The following result is due to Straubing [35].

Theorem 4.4 *If L is pure, 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 $\eta(x^k)$ is idempotent. By Proposition VII.2.2, it suffices to show that for every $u, v \in A^*$,

$$ux^kv \in L^* \text{ implies } ux^{k+1}v \in L^* \tag{4.1}$$

Suppose that $ux^kv \in L^*$. Then $ux^kv = u_1 \cdots u_n$, where each u_i belongs to $L \subseteq \{1\}$. 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 for some $r \in \{1, \ldots, n\}$, the r-th occurrence of x is not cut. Then $ux^{r-1} = u_1 \cdots u_{j-1} f$, $u_j = fx^t g$ and $x^q v = g u_{j+1} \cdots u_n$ for some $s, t \in A^*$ 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 $ux^{k+1} v = u_1 \cdots u_{j-1} f x^{t+1} g u_{j+1} \cdots u_n \in L^*$, proving (4.1) 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 $|x|$ factorizations 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 factorization $x = fg$, 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, $ux^{k+1} v = ux^{r-1} x x^s x x x^t v = (u^{r-1} f)(gx^s f)(gf)(gx^t v) \in L^*$, proving (4.1) in this case as well. □

Corollary 4.5 *If L is star-free and pure, then L^* is star-free.*

Proof. By Theorem VII.2.3, 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. □

4.3 Flower automata

Let L be a finite language if A^*. The *flower automaton* of L^* is the finite nodeterministic 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 recognizes L^*.

Example 4.2 Let $A = \{a, b\}$ and $L = \{a, ba, aab, aba\}$.

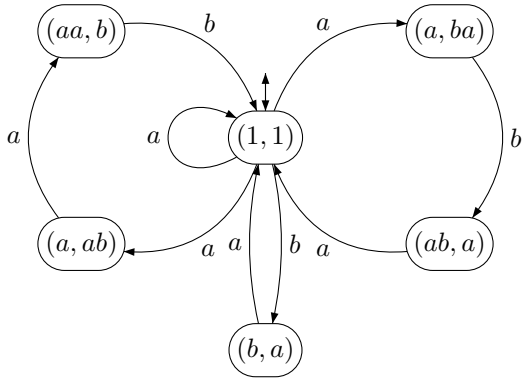


Figure 4.1. A flower automaton.

The transition monoid of the flower automaton of L^* is called the flower monoid of L^*. Since it recognizes L^*, the syntactic monoid of L^* is a quotient of it.

Recall that a subset X of the free monoid A^* is a code over A if for all $n, m > 0$ and $x_1, \ldots, x_n, x'_1, \ldots, x'_m \in X$, the condition

$$x_0 x_1 \cdots x_n = x'_1 x'_2 \cdots x'_m$$

implies $n = m$ and $x_i = x'_i$ for $1 \leqslant i \leqslant n$. In other words, a set X is a code if any word in X^+ can be written uniquely as a product of words in X.

Theorem 4.6 *Let X be a finite code. The natural morphism from the flower monoid of X^* onto its syntactic monoid is a locally trivial morphism.*

Proof. TO DO □

Chapter XI

Languages associated with DA

We denote by **DA** 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).

1 Algebraic characterizations of DA

2 Unambiguous star-free languages

Let A be a finite alphabet. The set of *unambiguous star-free* subsets 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.

a,

Proposition 2.1 *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. Further, any finite language is the disjoint union of the languages $\{u\}$, for $u \in F$. Thus every finite language is unambiguous star-free. □

Example 2.1 The language $\{a, c\}^* a\{1\}b\{b, c\}^*$ is unambiguous star-free (see Example X.4.1).

The aim of this section is to prove the following theorem

Theorem 2.2 *A language is unambiguous star-free if and only if its syntactic monoid is finite and belongs to* **DA**.

Proof. The easiest part of the proof relies on Theorem X.4.3. Let $L = L_0 a_1 L_1 \cdots a_k L_k$ be unambiguous marked product. Let M_0, ..., M_k and M be the respective syntactic monoids of L_0, ..., L_K and L.

Lemma 2.3 *If $M_0, \ldots, M_k$ belong to* **DA**, *so is M.*

Proof. TO DO □

Chapter XII

Semidirect product and wreath product

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 $T^1 \times S$ into S such that, for all $s, s_1, s_2 \in S$ and $t, t_1, t_2 \in T$,

(1) $(t_1 t_2) \cdot s = t_1(t_2 \cdot s)$

(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 $X \circ Y$, is the transformation semigroup $(P \times Q, W)$ where W consists of all pairs (f, t), with f is a function from Q into 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 $S \circ T$. 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 an action $(t,f) \mapsto t\cdot f$ of S^{T^1} on T by setting $t'\cdot(t\cdot f) = (t't)\cdot f$. 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 into $S^Q \times T$. 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 $\pi_1 \colon Q_1 \to P_1$ and $\pi_2 \colon Q_2 \to P_2$ be the surjective mappings defining the divisions. Let $\pi = \pi_1 \times \pi_2 \colon Q_1 \times Q_2 \to P_1 \times P_2$. 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 denote by c_p 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 into $X^1 \circ Y^1$.

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 start with commutative monoids.

Proposition 3.1 *Every commutative monoid divides 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 transform each element on N into the product of its coordinates. Then φ is clearly surjective and thus M is a quotient of N. □

We now study decompositions involving $\tilde{U}_n$ and U_n.

Proposition 3.2 *For every $n > 0$, U_n divides U_2^n and $\tilde{U}_n$ divides $\tilde{U}_2^n$.*

Proof. Arguing by induction on n, it suffices to verify that U_n divides $U_{n-1} \times U_2$. But a simple computation shows that U_n is isomorphic to the submonoid N of $U_{n-1} \times U_2$ defined as follows:

$$N = \{(1,1)\} \cup \{(a_i, a_1) \mid 0 \leqslant i \leqslant n-1\} \cup \{(a_1, a_2)\}$$

A dual proof works for $\tilde{U}_n$. □

A more precise result follows from Proposition **??**: 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 3.3 *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 $U_n \circ U_2$. □

It follows now immediately from Propositions 3.3 and 3.2:

Corollary 3.4 *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 **??**, these semigroups are left locally 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 3.5 *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 3.6 *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 3.6 and Lemma 3.5 now give immediately.

Corollary 3.7 *Every semigroup of* $\mathbf{D_n}$ *divides a wreath product of copies of* $\bar{\mathbf{2}}$.

The $\mathcal{R}$-trivial monoids admit also a simple decomposition.

Theorem 3.8 *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 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

$$\hat{s} = (f_{n-1}, \ldots, f_2, a_1)$$

with

$$a_1 = \begin{cases} 1 & \text{if } s = 1 \\ 0 & \text{if } s \neq 1 \end{cases}$$

where $f_{i+1} : \underbrace{U_1 \times \cdots \times U_1}_{i \text{ times}} \to U_1$ 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 3.9 *Let M be a finite aperiodic monoid and let $\pi : A^* \to M$ be a surjective morphism. Then one of the following cases occur:*

(1) *M is a monogenic monoid,*
(2) *M is isomorphic to $\tilde{U}_n$ for some $n > 0$,*
(3) *there is a proper partition $A = B \cup C$ such that $\pi((B^*C)^*)$ and $\pi(B^*)$ are proper submonoids of M.*

Proof. Let $S = M \setminus \{1\}$. Since M is aperiodic, S is a subsemigroup of M. Let L be an $\mathcal{L}$-class of S, maximal for the order $\leqslant_{\mathcal{J}}$. First assume that S is the semigroup generated by L. If $|L| = 1$, S is monogenic and so is M. Otherwise, $|L| > 1$ and by Proposition V.2.5, L is regular and thus consists of $\mathcal{L}$-equivalent idempotents. Now, Proposition V.1.2 shows that if e and f are $\mathcal{L}$-equivalent idempotents, then $ef = e$ and $fe = f$. It follows that L is a semigroup and hence $S = L$. In particular, we are in the second case.

Now assume that the semigroup generated by L is strictly contained in S and put

$$B = \{a \in A \mid \pi(a) \in L\} \quad \text{and} \quad C = A \setminus B$$

As $\pi(A)$ generates S and $\pi(B)$ does not, B is a strict subset of A, so that C is nonempty. We claim that $\pi(B^+)$ and $\pi(B^*C)^+$ are proper subsemigroups of S. For the first part of the claim, we observe that $\pi(B)$ is contained in L and thus $\pi(B^+)$ is contained in the subsemigroup of S generated by L, which is a proper subsemigroup of S. For the second part, we observe that every element of $\pi(B^*C)^+$ is $<_{\mathcal{J}}$-below L and thus $\pi(B^*C)^+$ is contained in $S \setminus L$. It follows now from the claim that we are in the third case of the proposition. □

Proposition 3.10 *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 $L^1 \circ \bar{N}$.*

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.

Let $m \in M$. If $m \in L$, we set $\hat{m} = (g, c_1)$, where $g : N \to L^1$ is defined by $g(n) = nm$ for all $n \in N$. Otherwise, if $m \notin L$, we set $\hat{m} = (f, m)$, where $f(n) = 1$ for all $n \in N$.

Let $(l, n) \in L^1 \times N$. Then

$$\varphi(l, n) \cdot m = (ln) \cdot m = lnm$$

Now, if $m \in L$,

$$(l,n)\cdot \hat{m} = (l,n)(g,c_1) = (l\cdot g(n),1) = (lnm,1)$$

and since L is a left ideal, $lnm \in L$. On the other hand, if $m \in N$,

$$(l,n)\cdot \hat{m} = (l,n)(f,m) = (l\cdot f(n),nm) = (l,nm)$$

and since N is a monoid, $nm \in N$. In both cases, $\varphi((l,n)\cdot \hat{m}) = lnm$. It follows that φ is a covering and thus M divides $L^1 \circ \overline{N}$. □

Theorem 3.11 *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. Consider the three cases given by Proposition 3.9. If M is monogenic, then it is $\mathcal{R}$-trivial, and the result follows from Theorem 3.8. If M is isomorphic to $\tilde{U}_n$ for some $n > 0$, the result follows from Corollary 3.4. Finally, suppose there is a proper partition $A = B \cup C$ such that $L = \pi((B^*C)^*)$ and $M = \pi(B^*)$ are proper submonoids of M. Then L is a left ideal of M, and since $A^* = (B^*C)^* \cup B^*$, $M = L \cup N$. Thus by Proposition 3.10, 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 and $\overline{N}$ also divide wreath products of copies of U_2, since $U_2 = U_2^1 = \overline{U}_2$. Finally, M itself divides a wreath product of copies of U_2. □

Proposition 3.12 *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 property $\mathcal{P}$, 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 admit without proof our last decomposition result (see the Notes section).

Proposition 3.13 *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 3.14 *Let $\pi : G \to H$ be a surjective morphism and let $K = \pi^{-1}(1)$. Then G is isomorphic to a subgroup of $K \circ H$.*

Proof. TO DO. □

4 Exercises.

4.1 Semidirect product and wreath product

Exercice 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.

Bibliography

[1] 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.

[2] J. Brzozowski, K. Culik and A. Gabrielan, Classification of noncounting events, *J. Comput. Syst. Sci.* **5** (1971), 41–53.

[3] A. H. Clifford and G. B. Preston, *The Algebraic Theory of Semigroups*, vol. 1, Amer. Math. Soc., 1961.

[4] A. H. Clifford and G. B. Preston, *The Algebraic Theory of Semigroups*, vol. 2, Amer. Math. Soc., 1967.

[5] J. H. Conway, *Regular Algebra and Finite Machines*, Chapman and Hall, London, 1971.

[6] A. De Luca and S. Varricchio, *Finiteness and Regularity in Semigroups and Formal Languages*, Springer-Verlag, 1999.

[7] 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.

[8] 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.

[9] R. Graham, On finite 0-simple semigroups and graph theory, *Math. Syst. Theory* **2** (1968), 325–339.

[10] P. A. Grillet, *Semigroups, An Introduction to the Structure Theory*, Marcel Dekker, Inc., New York, 1995.

[11] Higgins, *Techniques of Semigroup Theory*, World Scientific, 1999.

[12] J. M. Howie, *An Introduction to Semigroup Theory*, Academic Press, 1976.

[13] J. M. Howie, *Automata and Languages*, Clarendon Press, 1991.

[14] D. Krob, Complete sets of B-rational identities, *Theoret. Comp. Sci.* **89** (1991), 207–343.

[15] G. Lallement, *Semigroups and Combinatorial Applications*, Wiley and Sons, 1979.

[16] M. V. Lawson, *Finite Automata*, CRC Press, 2003.

[17] M. Lothaire, *Combinatorics on Words*, *Encyclopedia of Mathematics and its Applications* vol. 17, Cambridge University Press, 1983.

[18] M. Morse and G. A. Hedlund, Unending chess, symbolic dynamics and a problem in semigroups, *Duke Math. J.* **11** (1944), 1–7.

[19] M. Petrich, *Inverse Semigroups*, Wiley and Sons, 1984.

[20] J.-E. Pin, Hiérarchies de concaténation, *RAIRO Inform. Théor.* **18**,1 (1984), 23–46.

[21] J.-E. 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.

[22] J.-E. Pin, A property of the Schützenberger product, *Semigroup Forum* **35** (1987), 53–62.

[23] 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.

[24] J.-E. Pin, A variety theorem without complementation, *Russian Mathematics (Iz. VUZ)* **39** (1995), 80–90.

[25] J.-E. Pin, Syntactic semigroups, in *Handbook of formal languages, Vol. 1*, pp. 679–746, Springer, Berlin, 1997.

[26] J.-E. Pin, Algebraic tools for the concatenation product, *Theoretical Comput. Sci.* **292** (2003), 317–342.

[27] J.-E. Pin and P. Weil, Polynomial closure and unambiguous product, in *22th ICALP*, Berlin, 1995, pp. 348–359, *Lecture Notes in Comput. Sci.* n° 944, Springer.

[28] J.-E. Pin and P. Weil, A Reiterman theorem for pseudovarieties of finite first-order structures, *Algebra Universalis* **35**,4 (1996), 577–595.

[29] J.-E. Pin and P. Weil, Polynomial closure and unambiguous product, *Theory Comput. Systems* **30** (1997), 1–39.

[30] 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.".

[31] I. Simon, Factorization forests of finite height, *Theoret. Comput. Sci.* **72**,1 (1990), 65–94.

[32] H. Straubing, Aperiodic homomorphisms and the concatenation product of recognizable sets, *J. Pure Appl. Algebra* **15**,3 (1979), 319–327.

[33] H. Straubing, Families of recognizable sets corresponding to certain varieties of finite monoids, *J. Pure Appl. Algebra* **15**,3 (1979), 305–318.

[34] H. Straubing, A generalization of the Schützenberger product of finite monoids, *Theoret. Comput. Sci.* **13**,2 (1981), 137–150.

[35] H. Straubing, Relational morphisms and operations on recognizable sets, *RAIRO Inf. Theor.* **15** (1981), 149–159.

[36] A. Thue, Über unendliche Zeichenreihen, *Norske Vid. Selsk. Skr. I Math-Nat. Kl.* **7** (1906), 1–22.

[37] A. Thue, Über die gegenseitige Loge gleicher Teile gewisser Zeichenreihen, *Norske Vid. Selsk. Skr. I Math-Nat. Kl. Chris.* **1** (1912), 1–67.

Index