

AUTOMATIC SEQUENCES

Theory, Applications, Generalizations

JEAN-PAUL ALLOUCHE

CNRS, LSI, Orsay

JEFFREY SHALLIT

University of Waterloo



PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK
40 West 20th Street, New York, NY 10011-4211, USA
477 Williamstown Road, Port Melbourne, VIC 3207, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain
Dock House, The Waterfront, Cape Town 8001, South Africa
<http://www.cambridge.org>

© Jean-Paul Allouche Jeffrey Shallit 2003

This book is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without
the written permission of Cambridge University Press.

First published 2003

Printed in the United States of America

Typeface Times 11/14 pt. *System* L^AT_EX 2_ε [TB]

A catalog record for this book is available from the British Library.

Library of Congress Cataloging in Publication Data

Allouche, Jean-Paul, 1953-

Automatic sequences : theory, applications, generalizations / Jean-Paul Allouche, Jeffrey Shallit.

p. cm.

Includes bibliographical references and index.

ISBN 0-521-82332-3

1. Sequential machine theory. 2. Sequences (Mathematics) I. Shallit, Jeffrey Outlaw. II. Title.

QA267.5.S4 A55 2003

515'.24 – dc21 2002041262

ISBN 0 521 82332 3 hardback

Contents

<i>Preface</i>	<i>page</i> xiii
1 Stringology	1
1.1 Words	1
1.2 Topology and Measure	5
1.3 Languages and Regular Expressions	7
1.4 Morphisms	8
1.5 The Theorems of Lyndon and Schützenberger	10
1.6 Repetitions in Words	14
1.7 Overlap-Free Binary Words	16
1.8 Additional Topics on Repetitions	23
1.9 Exercises	24
1.10 Open Problems	30
1.11 Notes on Chapter 1	31
2 Number Theory and Algebra	39
2.1 Divisibility and Valuations	39
2.2 Rational and Irrational Numbers	39
2.3 Algebraic and Transcendental Numbers	41
2.4 Continued Fractions	44
2.5 Basics of Diophantine Approximation	48
2.6 The Three-Distance Theorem	53
2.7 Algebraic Structures	55
2.8 Vector Spaces	56
2.9 Fields	56
2.10 Polynomials, Rational Functions, and Formal Power Series	58
2.11 p -adic Numbers	62
2.12 Asymptotic Notation	63
2.13 Some Useful Estimates	63
2.14 Exercises	64
2.15 Open Problems	67
2.16 Notes on Chapter 2	67

3	Numeration Systems	70
3.1	Numeration Systems	70
3.2	Sums of Digits	74
3.3	Block Counting and Digital Sequences	77
3.4	Representation of Real Numbers	84
3.5	Sums of Sums of Digits	86
3.6	Base- k Representation with Alternate Digit Sets	101
3.7	Representations in Negative Bases	103
3.8	Fibonacci Representation	105
3.9	Ostrowski's α -Numeration System	106
3.10	Representations in Complex Bases	107
3.11	Exercises	112
3.12	Open Problems	118
3.13	Notes on Chapter 3	119
4	Finite Automata and Other Models of Computation	128
4.1	Finite Automata	128
4.2	Proving Languages Nonregular	136
4.3	Finite Automata with Output	137
4.4	Context-Free Grammars and Languages	143
4.5	Context-Sensitive Grammars and Languages	146
4.6	Turing Machines	146
4.7	Exercises	148
4.8	Open Problems	150
4.9	Notes on Chapter 4	150
5	Automatic Sequences	152
5.1	Automatic Sequences	152
5.2	Robustness of the Automatic Sequence Concept	157
5.3	Two-Sided Automatic Sequences	161
5.4	Basic Properties of Automatic Sequences	165
5.5	Nonautomatic Sequences	166
5.6	k -Automatic Sets	168
5.7	1-Automatic Sequences	169
5.8	Exercises	170
5.9	Open Problems	171
5.10	Notes on Chapter 5	171
6	Uniform Morphisms and Automatic Sequences	173
6.1	Fixed Points of Uniform Morphisms	173
6.2	The Thue–Morse Infinite Word	173
6.3	Cobham's Theorem	174
6.4	The Tower of Hanoi and Iterated Morphisms	177
6.5	Paperfolding and Continued Fractions	181
6.6	The k -Kernel	185

6.7	Cobham's Theorem for (k, l) -Numeration Systems	187
6.8	Basic Closure Properties	189
6.9	Uniform Transduction of Automatic Sequences	192
6.10	Sums of Digits, Polynomials, and Automatic Sequences	197
6.11	Exercises	201
6.12	Open Problems	207
6.13	Notes on Chapter 6	208
7	Morphic Sequences	212
7.1	The Infinite Fibonacci Word	212
7.2	Finite Fixed Points	213
7.3	Morphic Sequences and Infinite Fixed Points	215
7.4	Two-Sided Infinite Fixed Points	218
7.5	More on Infinite Fixed Points	226
7.6	Closure Properties	228
7.7	Morphic Images of Morphic Words	231
7.8	Locally Catenative Sequences	237
7.9	Transductions of Morphic Sequences	240
7.10	Exercises	242
7.11	Open Problems	244
7.12	Notes on Chapter 7	245
8	Frequency of Letters	247
8.1	Some Examples	247
8.2	The Incidence Matrix Associated with a Morphism	248
8.3	Some Results on Non-negative Matrices	249
8.4	Frequencies of Letters and Words in a Morphic Sequence	266
8.5	An Application	276
8.6	Gaps	278
8.7	Exercises	280
8.8	Open Problems	282
8.9	Notes	282
9	Characteristic Words	283
9.1	Definitions and Basic Properties	283
9.2	Geometric Interpretation of Characteristic Words	290
9.3	Application: Unusual Continued Fractions	291
9.4	Exercises	293
9.5	Open Problems	295
9.6	Notes on Chapter 9	295
10	Subwords	298
10.1	Introduction	298
10.2	Basic Properties of Subword Complexity	300
10.3	Results for Automatic Sequences	304
10.4	Subword Complexity for Morphic Words	306

10.5	Sturmian Words	312
10.6	Sturmian Words and k th-Power-Freeness	320
10.7	Subword Complexity of Finite Words	323
10.8	Recurrence	324
10.9	Uniform Recurrence	328
10.10	Appearance	333
10.11	Exercises	334
10.12	Open Problems	340
10.13	Notes on Chapter 10	340
11	Cobham's Theorem	345
11.1	Syndetic and Right Dense Sets	345
11.2	Proof of Cobham's Theorem	347
11.3	Exercises	350
11.4	Notes on Chapter 11	350
12	Formal Power Series	351
12.1	Examples	352
12.2	Christol's Theorem	354
12.3	First Application to Transcendence Results	359
12.4	Formal Laurent Power Series and Carlitz Functions	359
12.5	Transcendence of Values of the Carlitz–Goss Gamma Function	362
12.6	Application to Transcendence Proofs over $\mathbb{Q}(X)$	365
12.7	Furstenberg's Theorem	367
12.8	Exercises	371
12.9	Open Problems	375
12.10	Notes on Chapter 12	376
13	Automatic Real Numbers	379
13.1	Basic Properties of the Automatic Reals	379
13.2	Non-closure Properties of $L(k, b)$	382
13.3	Transcendence: An Ad Hoc Approach	385
13.4	Transcendence of the Thue–Morse Number	387
13.5	Transcendence of Morphic Real Numbers	391
13.6	Transcendence of Characteristic Real Numbers	393
13.7	The Thue–Morse Continued Fraction	394
13.8	Exercises	400
13.9	Open Problems	402
13.10	Notes on Chapter 13	403
14	Multidimensional Automatic Sequences	405
14.1	The Sierpiński Carpet	405
14.2	Formal Definitions and Basic Results	408
14.3	Subword Complexity	412
14.4	Formal Power Series	413
14.5	Automatic Sequences in Base $-1 + i$	414

14.6	The Pascal Triangle Modulo d	420
14.7	Exercises	424
14.8	Open Problems	425
14.9	Notes on Chapter 14	426
15	Automaticity	428
15.1	Basic Notions	428
15.2	Nondeterministic Automaticity	431
15.3	Unary Automaticity	433
15.4	Automaticity of Sequences	434
15.5	Exercises	436
15.6	Open Problems	436
15.7	Notes on Chapter 15	437
16	k -Regular Sequences	438
16.1	Basics	438
16.2	Robustness of the k -Regular Concept	441
16.3	Further Results	444
16.4	k -Regular Power Series	445
16.5	Additional Examples	447
16.6	Exercises	449
16.7	Open Problems	453
16.8	Notes on Chapter 16	454
17	Physics	455
17.1	The One-Dimensional Ising Model	457
17.2	The Rudin–Shapiro Sequence and the One-Dimensional Ising Model	459
17.3	Distribution Results for the Rudin–Shapiro Sequence	462
17.4	The One-Dimensional Schrödinger Operator	464
17.5	Exercises	466
17.6	Notes on Chapter 17	467
Appendix		
Hints, References, and Solutions for Selected Exercises		471
A.1	Chapter 1	471
A.2	Chapter 2	472
A.3	Chapter 3	472
A.4	Chapter 4	474
A.5	Chapter 5	474
A.6	Chapter 6	474
A.7	Chapter 7	475
A.8	Chapter 8	476
A.9	Chapter 9	476
A.10	Chapter 10	476
A.11	Chapter 11	477

A.12	Chapter 12	477
A.13	Chapter 13	477
A.14	Chapter 14	478
A.15	Chapter 15	478
A.16	Chapter 16	478
A.17	Chapter 17	479
	<i>Bibliography</i>	481
	<i>Index</i>	555

1

Stringology

In this chapter we introduce the basic objects of interest to this book: finite and infinite words. A set of words forms a language, a concept introduced in Section 1.3. Morphisms, discussed in Section 1.4, provide a way to transform words. Two of the basic theorems on words – the theorems of Lyndon and Schützenberger – are discussed in Section 1.5.

Repetitions in words are introduced in Section 1.6. Section 1.7 discusses the particular case of binary words avoiding a certain type of repetition called an overlap; this section is rather technical and can be omitted on a first reading. Finally, Section 1.8 briefly introduces some additional topics about repetitions.

1.1 Words

One of the fundamental mathematical objects we will study in this book is the *word*. A word is made up of *symbols* (or *letters*), which are usually displayed in a typewriter-like font like `this`. (We treat the notion of symbol as primitive and do not define it further.) Let Σ denote a nonempty set of symbols, or *alphabet*; in this book, Σ will almost always be finite. One alphabet is so important that we give it a special symbol: if k is an integer ≥ 2 , then we define

$$\Sigma_k = \{0, 1, \dots, k - 1\}.$$

Note that we sometimes identify symbols with the integers they represent, so that, for example,

$$\Sigma_2 = \{0, 1\} = \{0, 1\}.$$

We typically denote variables whose domain is Σ using the lowercase italic letters a, b, c, d . A *word* or *string* (we use the terms interchangeably) is a finite or infinite list of symbols chosen from Σ . Although we usually denote words by simply juxtaposing their symbols, such as `3245`, for clarity (particularly when negative integers are involved) we sometimes write them using an explicit concatenation operator, e.g., `Concat(3, 2, 4, 5)`. If unspecified, a word is assumed to be finite.

We typically use the lowercase italic letters s, t, u, v, w, x, y, z to represent finite words.

More precisely, let $[m..n]$ denote the set of integers $\{m, m + 1, \dots, n\}$. Then a finite word is a map from either $[0..n - 1]$ or $[1..n]$ to Σ . (The choice of the initial index gives us a little flexibility in defining words.) If $n = 0$, we get the *empty word*, which we denote by ϵ . The set of all finite words made up of letters chosen from Σ is denoted by Σ^* . For example, if $\Sigma = \{a, b\}$, then $\Sigma^* = \{\epsilon, a, b, aa, ab, ba, bb, aaa, \dots\}$. We let Σ^+ denote the set of all nonempty words over Σ .

If w is a finite word, then its *length* (the number of symbols it contains) is denoted by $|w|$. For example, if $w = \text{five}$, then $|w| = 4$. Note that $|\epsilon| = 0$. We can also count the occurrences of a particular letter in a word. If $a \in \Sigma$ and $w \in \Sigma^*$, then $|w|_a$ denotes the number of occurrences of a in w . Thus, for example, if $w = \text{abbab}$, then $|w|_a = 2$ and $|w|_b = 3$.

One of the fundamental operations on words is *concatenation*. We concatenate two finite words w and x by juxtaposing their symbols, and we denote this by wx . For example, if $w = \text{book}$ and $x = \text{case}$, then $wx = \text{bookcase}$. Concatenation of words is, in general, not commutative; for example, we have $xw = \text{casebook}$. However, concatenation is associative: we have $w(xy) = (wx)y$ for all words w, x, y . Notationally, concatenation is treated like multiplication, so that w^n denotes the word $www \dots w$ (n times). Note that the set Σ^* together with concatenation becomes an algebraic structure called a *monoid*, with the empty word ϵ playing the part of the identity element.

We say a word y is a *subword* or *factor* of a word w if there exist words x, z such that $w = xyz$. We say x is a *prefix* of w if there exists y such that $w = xy$. We say x is a *proper prefix* of w if $y \neq \epsilon$. We say that z is a *suffix* of w if there exists y such that $w = yz$. If $w = a_1a_2 \dots a_n$, then for $1 \leq i \leq n$, we define $w[i] = a_i$. If $1 \leq i \leq n$ and $i - 1 \leq j \leq n$, we define $w[i..j] = a_i a_{i+1} \dots a_j$. Note that $w[i..i] = a_i$ and $w[i..i - 1] = \epsilon$.

A *language* over Σ is a (finite or infinite) set of words – that is, a subset of Σ^* .

Example 1.1.1 The following are examples of languages:

$$\text{PRIMES2} = \{10, 11, 101, 111, \dots\}$$

(the primes expressed in base 2),

$$\text{EQ} = \{x \in \{0, 1\}^* : |x|_0 = |x|_1\}$$

(words containing an equal number of each symbol).

We now define infinite words (or infinite sequences – we use the terms interchangeably). We let \mathbb{Z} denote the integers, \mathbb{Z}^+ denote the positive integers, \mathbb{Z}^- denote the negative integers, and \mathbb{N} denote the non-negative integers. Then we will

usually take a *one-sided* (or *unidirectional*) right-infinite word $\mathbf{a} = a_0a_1a_2 \cdots$ to be a map from \mathbb{N} to Σ . We can form an infinite word by concatenating infinitely many finite words; for example,

$$\prod_{i \geq 1} w_i$$

denotes a word $w_1w_2w_3 \cdots$, which is infinite if and only if $w_i \neq \epsilon$ infinitely often.

Example 1.1.2 The following is an example of a right-infinite word:

$$\mathbf{q} = (q_n)_{n \geq 0} = 11001000010000001 \cdots,$$

where $q_n = 1$ if n is a square and 0 otherwise. The sequence \mathbf{q} is called the *characteristic sequence* of the perfect squares.

Sometimes, if subscripts become too cumbersome, we write $\mathbf{a} = a(0)a(1)a(2) \cdots$ instead. Also, instead of beginning indices at 0, occasionally we will use a map from \mathbb{Z}^+ to Σ , beginning our indices at 1, as the following example shows.

Example 1.1.3 Define

$$\mathbf{p} = (p_n)_{n \geq 1} = 0110101000101 \cdots,$$

the characteristic sequence of the prime numbers.

The set of all one-sided right-infinite words over Σ is denoted by Σ^ω . We define $\Sigma^\infty = \Sigma^* \cup \Sigma^\omega$.

A *left-infinite word* $\cdots a_{-3}a_{-2}a_{-1}$ is a map from \mathbb{Z}^- to Σ . The set of all left-infinite words is denoted by ${}^\omega\Sigma$.

A *two-sided* (or *bidirectional*) *infinite word* is a map from \mathbb{Z} to Σ . Such a word is of the form $\cdots c_{-2}c_{-1}c_0.c_1c_2c_3 \cdots$; the decimal point is a notational convention and not part of the word itself. We denote the set of all two-sided infinite words over Σ by $\Sigma^\mathbb{Z}$. In this book, infinite words are typically denoted in boldface. Unless otherwise indicated, infinite words are assumed to be one-sided and right-infinite.

We can produce one-sided infinite words from two-sided infinite words by ignoring the portion to the right or left of the decimal point. Suppose $\mathbf{w} = \cdots c_{-2}c_{-1}c_0.c_1c_2c_3 \cdots$. We define

$$\mathbf{L}(\mathbf{w}) = \cdots c_{-2}c_{-1}c_0,$$

a left-infinite word, and

$$\mathbf{R}(\mathbf{w}) = c_1c_2c_3 \cdots,$$

a right-infinite word.

The notions of subword, prefix, and suffix for finite words have evident analogues for infinite words. Let $\mathbf{w} = a_0a_1a_2 \cdots$ be an infinite word. For $i \geq 0$ we define $\mathbf{w}[i] = a_i$. For $i \geq 0$ and $j \geq i - 1$, we define $\mathbf{w}[i..j] = a_i a_{i+1} \cdots a_j$. We also define $\mathbf{w}[i..\infty] = a_i a_{i+1} \cdots$. If

$$\lim_{n \rightarrow \infty} \frac{|\mathbf{w}[0..n-1] |_b}{n}$$

exists and equals r , then the *frequency* of the symbol b in \mathbf{w} is defined to be r . We denote this frequency as $\text{Freq}_b(\mathbf{w})$.

Example 1.1.4 Consider the word \mathbf{q} from Example 1.1.2. Then $\text{Freq}_0(\mathbf{q}) = 1$ and $\text{Freq}_1(\mathbf{q}) = 0$.

Infinite words may be specified by the limit of a sequence of finite words. If w_1, w_2, w_3, \dots form a sequence of words with w_i a proper prefix of w_j for $i < j$, then $\lim_{n \rightarrow \infty} w_n$ is the unique infinite word of which w_1, w_2, \dots are all prefixes.

Let k be an integer ≥ 2 . A *k-aligned subword* of an infinite word $\mathbf{x} = a_0a_1a_2 \cdots$ is a subword of the form $a_{ki}a_{ki+1} \cdots a_{ki+k-1}$ for some integer $i \geq 0$.

We can also concatenate a finite word on the left with an infinite word on the right, but not vice versa. Clearly we cannot concatenate two right-infinite or two left-infinite words, but it is possible to concatenate a left-infinite word with a right-infinite word; see below. If x is a nonempty finite word, then x^ω is the right-infinite word $xxx \cdots$. Such a word is called *purely periodic*. An infinite word \mathbf{w} of the form $x y^\omega$ for $y \neq \epsilon$ is called *ultimately periodic*. If \mathbf{w} is ultimately periodic, then we can write it in the form $x y^\omega$ for finite words x, y with $y \neq \epsilon$. Then x is called a *preperiod* of \mathbf{w} , and y is called a *period*. (Sometimes we abuse terminology by calling the length $|x|$ the preperiod and $|y|$ the period.) If $|x|, |y|$ are chosen as small as possible, then x is called the *least preperiod*, and y is called the *least period*.

If L is a language, we define

$$L^\omega = \{w_1w_2w_3 \cdots : w_i \in L \setminus \{\epsilon\} \text{ for all } i \geq 1\}.$$

Thus, for example, Σ_2^ω is the set of all right-infinite words over $\{0, 1\}$. Similarly, we define

$${}^\omega L = \{\cdots w_{-2}w_{-1}w_0 : w_i \in L \setminus \{\epsilon\} \text{ for all } i \leq 0\}.$$

If w is a nonempty finite word, then by $w^{\mathbb{Z}}$ we mean the two-sided infinite word $\cdots www.www \cdots$. Using concatenation, we can join a left-infinite word $\mathbf{w} = \cdots c_{-2}c_{-1}c_0$ with a right-infinite word $\mathbf{x} = d_0d_1d_2 \cdots$ to form a new two-sided infinite word, as follows:

$$\mathbf{w}.\mathbf{x} := \cdots c_{-2}c_{-1}c_0.d_0d_1d_2 \cdots.$$

If L is a language, we define

$$L^{\mathbb{Z}} := \{\cdots w_{-2}w_{-1}w_0.w_1w_2 \cdots : w_i \in L \setminus \{\epsilon\} \text{ for all } i \in \mathbb{Z}\}.$$

If $w = a_1a_2 \cdots a_n$ and $x = b_1b_2 \cdots b_n$ are finite words of the same length, then by $w \text{ III } x$ we mean the word $a_1b_1a_2b_2 \cdots a_nb_n$, the *perfect shuffle* of w and x . For example, $\text{clip III aloe} = \text{calliope}$. A similar definition can be given for infinite words.

If $w = a_1a_2 \cdots a_n$ is a finite word, then by w^R we mean the *reversal* of the word w , that is, $w^R = a_na_{n-1} \cdots a_2a_1$. For example, $(\text{drawer})^R = \text{reward}$. Note that $(wx)^R = x^Rw^R$. A word w is a *palindrome* if $w = w^R$. Examples of palindromes in English include *deified*, *rotator*, *repaper*, and *redivider*.

If $\mathbf{w} = a_0a_1a_2 \cdots$ is a one-sided right-infinite word, then we define the *shift map* $\mathcal{S}(\mathbf{w})$ to be the word $a_1a_2a_3 \cdots$. Similarly, for $k \geq 0$, we have $\mathcal{S}^k(\mathbf{w}) = a_k a_{k+1} a_{k+2} \cdots$. For $k < 0$, we define $\mathcal{S}^k(\mathbf{w}) = u\mathbf{w}$ for an arbitrarily chosen word u of length k . For two-sided infinite words and $k \in \mathbb{Z}$, we define

$$\mathcal{S}^k(\cdots a_{-2}a_{-1}a_0.a_1a_2a_3 \cdots) = a_{k-2}a_{k-1}a_k.a_{k+1}a_{k+2}a_{k+3} \cdots$$

This notation is also extended to finite words, where for $k \geq 0$ we define

$$\mathcal{S}^k(a_0a_1 \cdots a_j) = \begin{cases} a_k a_{k+1} \cdots a_j & \text{if } 0 \leq k \leq j, \\ \epsilon, & \text{otherwise.} \end{cases}$$

If $\mathbf{w} = a_0a_1a_2 \cdots$ is an infinite word over Σ and $\mathbf{x} = b_0b_1b_2 \cdots$ is an infinite word over Δ , then by $\mathbf{w} \times \mathbf{x}$ we mean the infinite word $c_0c_1c_2 \cdots$ over $\Sigma \times \Delta$ defined by $c_i = (a_i, b_i)$. We also extend the notation \times to apply to finite words of the same length.

If Σ is an ordered set, we can define an ordering on words of Σ^ω . We define a lexicographic order Σ^ω as follows: let $\mathbf{w} = a_1a_2a_3 \cdots$ and $\mathbf{x} = b_1b_2b_3 \cdots$. Define $\mathbf{w} < \mathbf{x}$ if there exists an index $i \geq 0$ such that $a_j = b_j$ for $j \leq i$ and $a_{i+1} < b_{i+1}$. A similar definition can be given for finite words of the same length.

Let $\mathbf{w} = a_0a_1a_2 \cdots$ be an infinite word over Σ , and let k be an integer ≥ 1 . The *k-block compression* of \mathbf{w} , which we write as $\text{comp}(\mathbf{w}, k)$, is an infinite word $b_0b_1b_2 \cdots$ over the alphabet Σ^k defined by $b_i = (a_{ki}, a_{ki+1}, \dots, a_{ki+k-1})$.

If L_1, L_2 are languages over Σ , we define $L_1L_2 = \{xy : x \in L_1, y \in L_2\}$.

1.2 Topology and Measure

Let A be a set, and let \mathcal{T} be a collection of subsets of A . Recall that we say (A, \mathcal{T}) is a *topological space*, or just a *topology*, if

- (i) \emptyset and A are members of \mathcal{T} ;
- (ii) if $(X_i)_{i \in I}$ are members of \mathcal{T} , then so is $\bigcup_{i \in I} X_i$;
- (iii) if $(X_i)_{1 \leq i \leq n}$ are members of \mathcal{T} for some integer $n \geq 1$, then so is $\bigcap_{1 \leq i \leq n} X_i$.

The elements of \mathcal{T} are called *open sets*. A subset $F \subseteq A$ is called *closed* if its complement $A \setminus F$ is open. A topology may be specified by providing a *base* B ; this is a collection of open sets such that each element of \mathcal{T} may be expressed as a union of elements of B . A topology may be also specified by providing a *sub-base*

D of B ; this is a collection of open sets such that each element of B can be expressed as a nonempty finite intersection of elements of D .

Example 1.2.1 Let $A = \mathbb{R}$, and let the topology T be specified by letting B , a base, consist of the open intervals of the form (a, b) with $a, b \in \mathbb{R}$ and $a < b$.

Let Σ be a finite alphabet. We can specify a natural topology on Σ^ω , the set of one-sided right-infinite words over Σ , by specifying a sub-base D as follows:

$$D = \bigcup_{\substack{j \geq 0 \\ a \in \Sigma}} D_{j,a},$$

where $D_{j,a}$ consists of those words \mathbf{w} such that $\mathbf{w}[j] = a$. Base elements, which are nonempty finite intersections of the $D_{j,a}$, are of the form $\Sigma^{i_1} a_1 \Sigma^{i_2} a_2 \cdots \Sigma^{i_j} a_j \Sigma^\omega$, where $j, i_1, i_2, \dots, i_j \geq 0$ are integers and $a_1, a_2, \dots, a_j \in \Sigma$. Such a set is called a *cylinder*.

Theorem 1.2.2 *The open sets in Σ^ω are precisely those sets of the form $L\Sigma^\omega$, with $L \subseteq \Sigma^*$.*

Proof. Since by definition the $D_{j,a}$ form a sub-base for the topology, every base element is of the form $\Sigma^{i_1} a_1 \Sigma^{i_2} a_2 \cdots \Sigma^{i_j} a_j \Sigma^\omega$, where $j, i_1, i_2, \dots, i_j \geq 0$ are integers and $a_1, a_2, \dots, a_j \in \Sigma$. Thus every base element is of the form $L\Sigma^\omega$, where $L = \Sigma^{i_1} a_1 \Sigma^{i_2} a_2 \cdots \Sigma^{i_j} a_j$. Now by definition each open set is a union of sets of the form $L_i \Sigma^\omega$. But $\bigcup_{i \in I} L_i \Sigma^\omega = (\bigcup_{i \in I} L_i) \Sigma^\omega$.

For the converse, we need to show that $L\Sigma^\omega$ is open. But $L\Sigma^\omega = \bigcup_{x \in L} x \Sigma^\omega$, and each element of the form $x \Sigma^\omega$ for $x \in \Sigma^*$ is clearly a base element. ■

Let (X, T) be a topological space, and $A \subseteq X$. We say that $x \in X$ is a *limit point* of A if every open set containing x intersects $A \setminus \{x\}$. The set of all limit points of A is called the *derived set* and is sometimes written as A' . If $A = A'$, we say A is *perfect*.

Recall that a *metric* on a set A is a map $d : A \rightarrow \mathbb{R}^{\geq 0}$ such that

- (i) for $x, y \in A$ we have $d(x, y) = 0$ if and only if $x = y$;
- (ii) for $x, y \in A$ we have $d(x, y) = d(y, x)$;
- (iii) for $x, y, z \in A$ we have the *triangle inequality* $d(x, z) \leq d(x, y) + d(y, z)$.

Here by $\mathbb{R}^{\geq 0}$ we mean the non-negative real numbers. The pair (A, d) is called a *metric space*.

A metric d induces a topology as follows: we take as a base the family of all open balls of the form $\{x \in A : d(x, y) < r\}$ for $y \in A$ and $r > 0$.

We can make Σ^ω into a metric space by defining

$$d(\mathbf{x}, \mathbf{y}) = \begin{cases} 0 & \text{if } \mathbf{x} = \mathbf{y}, \\ 2^{-n} & \text{otherwise,} \end{cases}$$

where $n = \min\{i : \mathbf{x}[i] \neq \mathbf{y}[i]\}$. Intuitively, two infinite sequences are “close together” if they agree on a long prefix. Note that in addition to the triangle inequality, d satisfies the stronger *ultrametric inequality*

$$d(\mathbf{x}, \mathbf{y}) \leq \max(d(\mathbf{x}, \mathbf{z}), d(\mathbf{y}, \mathbf{z}))$$

for all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \Sigma^\omega$. It is not difficult to see that the topology induced by d is the same as the topology mentioned above.

The *closure* of a set $X \subseteq \Sigma^\omega$ is defined to be the intersection of all closed subsets of Σ^ω containing X ; it is denoted by $\text{Cl}(X)$. Alternatively, $\mathbf{w} \in \text{Cl}(X)$ if for all real $\delta > 0$ there exists $\mathbf{x} \in X$ such that $d(\mathbf{w}, \mathbf{x}) < \delta$.

Theorem 1.2.3 *Let $X \subseteq \Sigma^\omega$, and let $\mathbf{w} \in \Sigma^\omega$. Then $\mathbf{w} \in \text{Cl}(X)$ if and only if every prefix of \mathbf{w} is the prefix of some word in X .*

Proof. We have $\mathbf{w} \in \text{Cl}(X)$ if and only if for all $k \geq 0$ there exists $\mathbf{x} \in X$ with $d(\mathbf{w}, \mathbf{x}) \leq 2^{-k}$, if and only if for all $k \geq 0$ there exists $\mathbf{x} \in X$ which agrees with \mathbf{w} on the first k terms. ■

We can extend the metric d to Σ^∞ by introducing a new symbol b , not in Σ , and identifying each finite word w with the right-infinite word $wb^\omega \in (\Sigma \cup \{b\})^\omega$.

Finally, we can put a measure m on Σ^ω by defining the measure of the cylinders

$$m(\Sigma^{i_1} a_1 \Sigma^{i_2} a_2 \cdots \Sigma^{i_j} a_j \Sigma^\omega) = k^{-j},$$

where $k = \text{Card } \Sigma$.

1.3 Languages and Regular Expressions

As we have seen above, a language over Σ is a subset of Σ^* . Languages may be of finite or infinite cardinality. We start by defining some common operations on languages.

Let $L, L_1, L_2 \subseteq \Sigma^*$ be languages. Recall that we define the *product of languages* by

$$L_1 L_2 = \{wx : w \in L_1, x \in L_2\}.$$

We define $L^0 = \{\epsilon\}$ and define L^i as LL^{i-1} for $i \geq 1$. We define

$$L^{\leq i} = L^0 \cup L^1 \cup \cdots \cup L^i.$$

We define L^* as $\bigcup_{i \geq 0} L^i$; the operation L^* is sometimes called *Kleene closure*. We define $L^+ = L L^*$; the operation $+$ in the superscript is sometimes called *positive*

closure. If L is a language, then the *reversed* language is defined as follows: $L^R = \{x^R : x \in L\}$. Finally, we define the *quotient* of languages as follows:

$$L_1/L_2 = \{x \in \Sigma^* : \exists y \in L_2 \text{ such that } xy \in L_1\}.$$

We now turn to a common notation for representing some kinds of languages. A *regular expression* over the alphabet Σ is a well-formed word over the alphabet

$$\Sigma \cup \{\epsilon, \emptyset, (,), +, *\}.$$

(Exercise 64 makes this more precise.) In evaluating such an expression, $*$ represents Kleene closure and has highest precedence. Concatenation is represented by juxtaposition, and has next highest precedence. Finally, $+$ represents union and has lowest precedence. Parentheses are used for grouping.

If the word u is a regular expression, then $L(u)$ represents the language that u specifies. For example, consider the regular expression $u = (0+10)^*(1+\epsilon)$. Then $L(u)$ represents all finite words of 0's and 1's that do not contain two consecutive 1's. Frequently we will abuse the notation by referring to the language as the naked regular expression without the surrounding $L()$. A language L is said to be *regular* if $L = L(u)$ for some regular expression u .

Theorem 1.3.1 *Every finite language is regular.*

Proof. If $L = \{w_1, w_2, \dots, w_i\}$, then a regular expression for L is just $w_1 + w_2 + \dots + w_i$. ■

1.4 Morphisms

In this section we introduce a fundamental tool of formal languages, the *homomorphism*, or just *morphism* for short. Let Σ and Δ be alphabets. A morphism is a map h from Σ^* to Δ^* that obeys the identity $h(xy) = h(x)h(y)$ for all words $x, y \in \Sigma^*$. Typically, we use the Latin letters f, g, h and Greek letters $\varphi, \theta, \mu, \sigma, \rho$ to denote morphisms.

Clearly if h is a morphism, then we must have $h(\epsilon) = \epsilon$. Furthermore, once h is defined for all elements of Σ , it can be uniquely extended to a map from Σ^* to Δ^* . Henceforth, when we define a morphism, we will always give it by specifying its action on Σ .

Example 1.4.1 Let $\Sigma = \{e, m, o, s\}$, let $\Delta = \{a, e, l, n, r, s, t\}$, and define

$$\begin{aligned} h(m) &= \text{ant}, \\ h(o) &= \epsilon, \\ h(s) &= \text{ler}, \\ h(e) &= s. \end{aligned}$$

Then $h(\text{moose}) = \text{antlers}$.

If $\Sigma = \Delta$, then we can iterate the application of h . We define $h^0(a) = a$ and $h^i(a) = h(h^{i-1}(a))$ for all $a \in \Sigma$.

Example 1.4.2 Let $\Sigma = \Delta = \{0, 1\}$. Define the *Thue–Morse morphism* $\mu(0) = 01$ and $\mu(1) = 10$. Then $\mu^2(0) = 0110$ and $\mu^3(0) = 01101001$.

There are various parameters associated with a morphism $h : \Sigma^* \rightarrow \Delta^*$. We define $\text{Width } h = \max_{a \in \Sigma} |h(a)|$, $\text{Depth } h = \text{Card } \Sigma$, and $\text{Size } h = \sum_{a \in \Sigma} |h(a)|$.

We can classify morphisms into different groups, as follows: if there is a constant k such that $|h(a)| = k$ for all $a \in \Sigma$, then we say that h is *k-uniform* (or just *uniform*, if k is clear from the context). A 1-uniform morphism is called a *coding*. We typically use the Greek letters τ and ρ to denote codings. A morphism is said to be *expanding* if $|h(a)| \geq 2$ for all $a \in \Sigma$.

If $h(a) \neq \epsilon$ for all $a \in \Sigma$, then h is *nonerasing*. If $h(a) = \epsilon$ for all $a \in \Sigma$, then we say h is *trivial*. If there exists an integer $j \geq 1$ such that $h^j(a) = \epsilon$, then the letter a is said to be *mortal*. The set of mortal letters associated with a morphism h is denoted by M_h . The *mortality exponent* of a morphism h is defined to be the least integer $t \geq 0$ such that $h^t(a) = \epsilon$ for all $a \in M_h$. (If $M_h = \emptyset$, we take $t = 0$.) We write the mortality exponent as $\text{exp}(h) = t$. It is easy to prove that $\text{exp}(h) \leq \text{Card } M_h$; see Exercise 3.

We also define the notion of *inverse homomorphism* of languages. Given $h : \Sigma^* \rightarrow \Delta^*$ and a language L , we define

$$h^{-1}(L) = \{x \in \Sigma^* : h(x) \in L\}.$$

We can also apply morphisms to infinite words. If $\mathbf{w} = c_0c_1c_2 \cdots$ is a right-infinite word, then we define

$$h(\mathbf{w}) = h(c_0)h(c_1)h(c_2) \cdots.$$

If $\mathbf{w} = \cdots c_{-2}c_{-1}c_0.c_1c_2 \cdots$ is a two-sided infinite word, and h is a morphism, then we define

$$h(\mathbf{w}) := \cdots h(c_{-2})h(c_{-1})h(c_0).h(c_1)h(c_2) \cdots. \quad (1.1)$$

We now introduce the notion of a *primitive morphism*. A morphism $h : \Sigma^* \rightarrow \Sigma^*$ is said to be *primitive* if there exists an integer $n \geq 1$ such that for all $a, b \in \Sigma$, a occurs in $h^n(b)$.

One reason why primitive morphisms are of interest is that if h is primitive, then the growth rate of $|h^r(a)|$ is essentially independent of a . We have the following

Theorem 1.4.3 *Let $h : \Sigma^* \rightarrow \Sigma^*$ be a primitive morphism. Then there exists a constant C (which does not depend on n but may depend on $\text{Width } h$ and $\text{Depth } h$) such that $|h^n(b)| \leq C|h^n(c)|$ for all $b, c \in \Sigma$ and all $n \geq 0$.*

Proof. Let $W = \text{Width } h$. Since h is primitive, there exists an integer $e \geq 1$ such that for all $b, c \in \Sigma$ we have $h^e(c) \in \Sigma^* b \Sigma^*$. Thus for $r \geq 1$ we have

$$\begin{aligned} |h^{er}(c)| &= |h^{e(r-1)}(h^e(c))| \\ &= |h^{e(r-1)}(xby)| \quad \text{for some } x, y \in \Sigma^* \\ &\geq |h^{e(r-1)}(b)|. \end{aligned}$$

Also $|h^{er}(b)| = |h^e(h^{e(r-1)}(b))| \leq W^e |h^{e(r-1)}(b)|$. Putting these bounds together, we get $|h^{er}(b)| \leq W^e |h^{er}(c)|$.

Now write $n = er + i$ for some $r \geq 0$ and $0 \leq i < e$. If $r = 0$, we have $|h^i(b)| \leq W^i \leq W^i |h^i(c)|$. If $r \geq 1$, then

$$\begin{aligned} |h^{er+i}(b)| &\leq W^i |h^{er}(b)| \\ &\leq W^{i+e} |h^{er}(c)| \\ &\leq W^{i+e} |h^{er+i}(c)|. \end{aligned}$$

It follows that $|h^n(b)| \leq W^{2e-1} |h^n(c)|$, so we may take $C = W^{2e-1}$. ■

Exercise 8.8 explores how big e can be. Note: Theorem 1.4.3 is made more precise in Proposition 8.4.1.

Let $h : \Sigma^* \rightarrow \Sigma^*$ be a morphism. A finite or infinite word w such that $h(w) = w$ is said to be a *fixed point* of h . If there exists a letter $a \in \Sigma$ such that $h(a) = ax$, and $x \notin M_h^*$, we say h is *prolongable* on a . In this case, the sequence of words $a, h(a), h^2(a), \dots$ converges, in the limit, to the infinite word

$$h^\omega(a) := a x h(x) h^2(x) \cdots,$$

which is a fixed point of h , that is, $h(h^\omega(a)) = h^\omega(a)$. Furthermore, it is easy to see that $h^\omega(a)$ is the unique fixed point of h which starts with a . If $\mathbf{w} = h^\omega(a)$, then we call \mathbf{w} a *pure morphic sequence*. If there is a coding $\tau : \Sigma \rightarrow \Delta$ and $\mathbf{w} = \tau(h^\omega(a))$, then we call \mathbf{w} a *morphic sequence*.

1.5 The Theorems of Lyndon and Schützenberger

In this section, we prove two beautiful and fundamental theorems due to Lyndon and Schützenberger. We start with one of the simplest and most basic results on words, sometimes known as Levi's lemma:

Theorem 1.5.1 *Let $u, v, x, y \in \Sigma^*$, and suppose that $uv = xy$. If $|u| \geq |x|$, there exists $t \in \Sigma^*$ such that $u = xt$ and $y = tv$. If $|u| < |x|$, there exists $t \in \Sigma^+$ such that $x = ut$ and $v = ty$.*

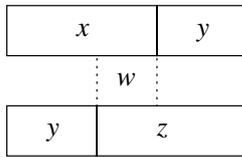
Proof. Left to the reader. ■

Now we can state the first theorem of Lyndon and Schützenberger:

Theorem 1.5.2 *Let $y \in \Sigma^*$ and $x, z \in \Sigma^+$. Then $xy = yz$ if and only if there exist $u, v \in \Sigma^*$ and an integer $e \geq 0$ such that $x = uv$, $z = vu$, and $y = (uv)^e u = u(vu)^e$.*

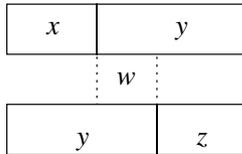
Proof. \implies : The proof is by induction on $|y|$. If $|y| = 0$, then we can take $v = x = z$, $u = \epsilon$, and $e = 0$. Thus suppose that $|y| \geq 1$. There are two cases.

Case I: If $|x| \geq |y|$, then we have a situation like the following:



By Levi's lemma there exists $w \in \Sigma^*$ such that $x = yw$ and $z = wy$. Now take $u = y$, $v = w$, $e = 0$, and we are done.

Case II: Now suppose that $|x| < |y|$. Then we have a situation like the following:



By Levi's lemma there exists $w \in \Sigma^+$ such that $y = wz = xw$. By induction (since $|w| = |y| - |z| < |y|$), we know there exist u, v, e such that

$$\begin{aligned} x &= uv, \\ z &= vu, \\ w &= (uv)^e u = u(vu)^e, \end{aligned}$$

so it follows that $y = u(vu)^{e+1} = (uv)^{e+1}u$.

\longleftarrow : We have

$$\begin{aligned} xy &= uv(uv)^e u = (uv)^{e+1}u, \\ yz &= u(vu)^e vu = u(vu)^{e+1}, \end{aligned}$$

and these words are identical. ■

We now state and prove the second theorem of Lyndon and Schützenberger.

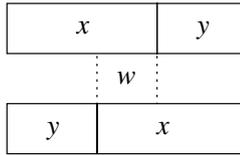
Theorem 1.5.3 *Let $x, y \in \Sigma^+$. Then the following three conditions are equivalent:*

- (1) $xy = yx$.
- (2) *There exist integers $i, j > 0$ such that $x^i = y^j$.*
- (3) *There exist $z \in \Sigma^+$ and integers $k, l > 0$ such that $x = z^k$ and $y = z^l$.*

Proof. We show that (1) \implies (3), (3) \implies (2), and (2) \implies (1).

(1) \implies (3): By induction on $|xy|$. If $|xy| = 2$, then $|x| = |y| = 1$, so $x = y$ and we may take $z = x = y, k = l = 1$.

Now assume the implication is true for all x, y with $|xy| < n$. We prove it for $|xy| = n$. Without loss of generality, assume $|x| \geq |y|$. Then we have a situation like the following:



Hence there exists $w \in \Sigma^*$ such that $x = wy = yw$. If $|w| = 0$ then $x = y$ and we can take $z = x = y, k = l = 1$.

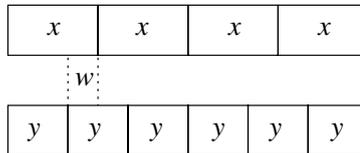
Otherwise $|w| \geq 1$. We have $|wy| = |x| < |xy| = n$, so the induction hypothesis applies, and there exists $z \in \Sigma^+$ and integers $k, l > 0$ such that $w = z^k, y = z^l$. It follows that $x = wy = z^{k+l}$.

(3) \implies (2): By (3) there exist $z \in \Sigma^+$ and integers $k, l > 0$ such that $x = z^k$ and $y = z^l$. Hence, taking $i = l, j = k$, we get

$$x^i = (z^k)^i = z^{kl} = (z^l)^k = (z^l)^j = y^j,$$

as desired.

(2) \implies (1): We have $x^i = y^j$. If $|x| = |y|$ then we must have $i = j$ and so $x = y$. Otherwise, without loss of generality assume $|x| > |y|$. Then we have a situation like the following:



That is, there exists $w \in \Sigma^+$ such that $x = yw$. Hence $x^i = (yw)^i = y^i w^i$, and so $y(yw)^{i-1} w = y^j$. Therefore $(yw)^{i-1} w = y^{j-1}$, and so, by multiplying by y on the right, we get $(yw)^i = y^j$. Hence $(yw)^i = (yw)^j$, and hence $yw = wy$. It follows that $x = yw = wy$ and $xy = (yw)y = y(yw) = yx$. ■