Combinatorics, Automata and Number Theory

CANT

Edited by

Valérie Berthé

LIRMM - Université Montpelier II - CNRS UMR 5506
161 rue Ada, F-34392 Montpellier Cedex 5, France

Michel Rigo

Université de Liège, Institut de Mathématiques
Grande Traverse 12 (B 37), B-4000 Liège, Belgium

Contents

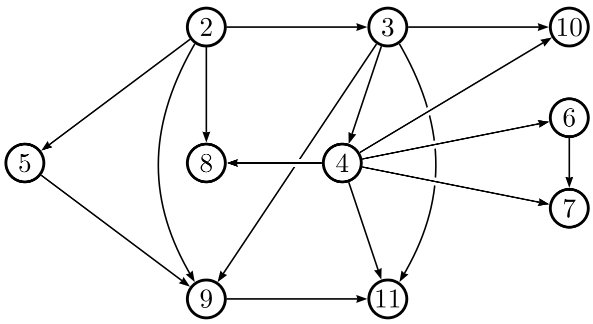
Introduction

As the title may suggest, this book is about *combinatorics on words*, *automata and formal language theory*, as well as *number theory*. This collaborative work gives a glimpse of the active community working in these interconnected and even intertwined areas. It presents several important tools and concepts usually encountered in the literature and it reveals some of the exciting and non-trivial relationships existing between the considered fields of research. This book is mainly intended for graduate students or research mathematicians and computer scientists interested in combinatorics on words, theory of computation, number theory, dynamical systems, ergodic theory, fractals, tilings, and stringology. We hope that some of the chapters can serve as useful material for lecturing at a master level.

The outline of this project has germinated after a very successful international eponymous school organised at the University of Liège (Belgium) in 2006 and supported by the European Union with the help of the European Mathematical Society (EMS). Parts of a preliminary version of this book were used as lecture notes for the second edition of the school organised in June 2009 and mainly supported by the European Science Foundation (ESF) through the AutoMathA programme. For both events, we acknowledge also financial support from the University of Liège and the Belgian funds for scientific research (FNRS).

We have selected ten topics which are directed towards the fundamental three directions of this project (namely, combinatorics, automata and number theory) and they naturally extend to dynamical systems and ergodic theory (see Chapters 6 and 7), but also to fractals and tilings (see Chapter 5) and spectral properties of matrices (see Chapter 11). Indeed, as it will be shown in particular in Chapter 7 there exist tight and fruitful links between properties sought for in dynamical systems and combinatorial properties of the corresponding words and languages. On the other hand, linear algebra and extremal matrix products are important tools in the framework of this book: some matrices are canonically associated with morphisms and graphs and a notion like joint spectral radius introduced in (Rota and Strang 1960) has therefore applications in automata theory or combinatorics on words.

Each chapter is intended to be self-contained and relies mostly on the introductory Chapter 1 presenting some preliminaries and general notions. Some of the major links existing between the chapters are given in the figure below.



Fifteen authors were collaborating on this volume. Most of them kindly served as lecturers for the CANT schools.

Let us succinctly sketch the general landscape without any attempt of being exhaustive. Short abstracts of each chapter are given below.

Combinatorics on words is a quite recent topics in (discrete) mathematics, and in the category of "Concrete Mathematics" according to the terminology introduced by (Graham, Knuth, and Patashnik 1989). It deals with problems that can be stated in a non-commutative monoid such as estimates on the factor complexity function for infinite words, construction and properties of infinite words, the study of unavoidable regularities or patterns, substitutive words, *etc.* In the spirit of Lothaire's seminal book series, see (Lothaire 1983), (Lothaire 2002) and (Lothaire 2005), but with a different focus put on interactions between fields of research, we will deal in this book with the complexity function counting factors occurring in an infinite word, properties and generalisations of automatic sequences in the sense of (Allouche and Shallit 2003) and also the equality problem for substitutive (or also called morphic) words, see Chapters 3, 4, and 10. Motivations to study words and their properties are coming, for instance, from the coding of orbits and trajectories by words. This constitutes the basis of symbolic dynamical systems (Lind and Marcus 1995). This explains why dynamical systems enter the picture, mainly in Chapter 6 and 7, and are at the origin of the introduction of the fractals studied in Chapter 5. A historical example is the study by M. Morse of recurrent geodesics on a surface with negative curvature (Morse 1921). As another example, similar ideas are found in connection with the word problem in group theory (Epstein, Cannon, Holt, et al. 1992). Moreover the use of combinatorics is sought in the analysis of algorithms, initiated by D. E. Knuth, and which greatly relies on number theory, asymptotic methods and computer algebra (Lothaire 2005), (Greene and Knuth 1990), (Knuth 2000). Reader interested in asymptotics methods and limiting properties of digital functions should in particular read Chapter 9.

Keep in mind that both combinatorics on words and theory of formal languages have important applications and interactions in computer science (Perrin and Pin 2003) and physics. To cite just a few: study and models of quasi-crystals, aperiodic order and quasiperiodic tilings, bio-informatics and DNA analysis, theory of parsing, algorithmic verification of large systems, coding theory, discrete geometry and more precisely discretisation for computer graphics on a raster display, *etc.* This shows that algorithmic issues have also an important role to play.

Two chapters of this book, Chapters 2 and 3, are dealing with **numeration systems**. Such systems provide a main bridge between number theory on the one hand, and words combinatorics and formal language theory on the other hand. Indeed any integer can be represented in a given numeration system, like the classical integer base q numeration system, as a finite word over a finite alphabet of digits $\{0, \dots, q-1\}$. This simple observation leads to the study of the relationships that can exist between the arithmetical properties of the integers and the syntactical properties of the corresponding representations. One of the deepest and most beautiful results in this direction is given by the celebrated theorem of Cobham (Cobham 1969) showing that the recognisability of a set of integers depends on the considered numeration system. This result can therefore be considered as one of the starting point of many investigations, for the last thirty years, about recognisable sets of integers and about non-standard or exotic numeration systems. Surprisingly, a recent extension of Cobham's theorem to the complex numbers leads to the famous Four Exponentials Conjecture (Hansel and Safer 2003). This is just one example of the fruitful relationship between formal language theory (including the theory of automata) and number theory. Many such examples will be presented here.

Numeration systems are not restricted to the representation of integers. They can also be used to represent real numbers with infinite words. One can think of continued fractions, integer or rational base representations, beta-expansions, *etc.* Again it is remarkable that some syntactical properties of the representations of reals may reflect number-theoretical properties, like transcendence, of the represented numbers. These questions are also treated in this book, see in particular Chapter 2 and Chapter 8. About Diophantine analysis or approximations of real numbers by algebraic numbers, striking developments through a fruitful interplay between Diophantine approximation and combinatorics on words can be observed, see again Chapter 8. Analogously, a rich source of challenging problems in analytic number theory comes from the study of digital functions, *i.e.*, functions defined in a way that depends on the digits in some numeration system. They are the object of Chapter 9.

The study of simple algorithmic constructions and transformations of infinite words plays here an important role. We focus in particular on the notion of **morphic words**, also called **substitutive words**. They are obtained iteratively by replacing letters with finite words. These words, as well as their associated symbolic dynamical systems, present a very rich behaviour. They occur in most of the chapters, see in particular Chapters 3, 4, 5, 6, 8 and 10. In the case where we replace letters with words of the same length, we obtain so-called automatic sequences. Several variations around the notion of morphic words are presented, as D0L systems (see Chapter 10), or else as adic words and transformations, and linearly recurrent subshifts. They occur in particular in Chapter 6 and Chapter 7. Note that most of the symbolic dynamical systems considered are of zero entropy, such as substitutive dynamical systems, odometers (see Chapter 6 and 9) or linearly recurrent systems (see Chapter 6).

Graphs and automata appear to be a very natural and powerful tool in this context. This is illustrated *e.g.* in Chapter 2 with special focus on operations performed on expansions of numbers realised by automata or transducers, or in Chapter 5 which is devoted to tilings by fractals whose boundary is described in terms of graphs. Graphs associated with substitutions appear ubiquitously, for instance, under the form of prefix-suffix graphs, of Rauzy graphs of words, or of the automata generating automatic sequences. Incidence matrices of graphs play here also an important role, hence the recurrence of notions like the spectral radius and its generalisations (see Chapter 11), or the importance of Perron–Frobenius' Theorem.

We are very pleased that Cambridge University Press proposed to consider this book, as the Lothaire's books, as part of the *Encyclopedia of Mathematics and its Applications* series.

Acknowledgements

We would like to express our gratitude to Shabnam Akhtari, Nicolas Bédaride, Emilie Charlier, Timo Jolivet, Tomi Kärki, Sébastien Labbé, Marion Le Gonidec, Geneviève Paquin, N. Pytheas Fogg who were kind enough to read drafts of this book. They pointed out mathematical and stylistic errors and suggested many improvements. Also we would like to thank our former editor Peter Thompson and our present editor Clare Dennison whose constant support has been a precious help for us all along this project, and Pierre Lecomte for his invaluable contribution in the organisation of the CANT schools.

Paris, November 2009
V. Berthé and M. Rigo

Let us present the different contributions for this book.

Chapter 2 by Ch. Frougny and J. Sakarovitch
Number representation and finite automata

In this chapter, numbers are represented by their expansion in a base, or more generally, with respect to a basis, hence by words (finite or infinite) over an alphabet of digits.

Is the set of expansions for all integers or all reals (within an interval) recognised by a finite automaton? Which operations on numbers translate into functions on number expansions that are realised by finite transducers? These are some of the questions that are treated in this chapter. The classical representation in an integer base is first considered, then the representation in a real base and in some associated basis. Finally, representations in canonical number systems and in rational bases are briefly studied.

Chapter 3 by P. Lecomte and M. Rigo
Abstract numeration systems

The motivation for the introduction of abstract numeration systems stems from the celebrated theorem of Cobham dating back to 1969 about the so-called recognisable sets of integers in any integer base numeration system. An abstract numeration system is simply an infinite genealogically ordered (regular) language. In particular, this notion extends the usual integer base numeration systems as well as more elaborated numeration systems such as those based on a Pisot number. In this general setting, we study in details recognisable sets of integers, *i.e.*, the corresponding representations are accepted by a finite automaton. The main theme is the link existing between the arithmetic properties of integers and the syntactical properties of the corresponding representations in a given numeration system. Relationship with automatic sequences and substitutive words is also investigated, providing an analogue to another famous result of Cobham from 1972 about k-automatic sequences. Finally, the chapter ends with the representation of real numbers in an abstract numeration system.

Chapter 4 by J. Cassaigne and F. Nicolas
Factor complexity

The factor complexity function $p(n)$ of an infinite word is studied thoroughly. Tools such as special factors and Rauzy graphs are introduced, then applied to several problems, including practical computation of the factor complexity of various kinds of words, or the construction of words having a complexity asymptotically equivalent to a specified function.

This chapter includes a complete proof of Pansiot's characterisation of the complexity function of purely morphic words, and a proof of a conjecture of Heinis on the limit $p(n)/n$.

The authors would like to thank Jean-Paul Allouche for his bibliographic help, Juhani Karhumäki for his kind hospitality during the redaction of this chapter, and Christian Mauduit for his participation in the proof of Theorem 4.7.15. F. Nicolas was supported by the Academy of Finland under the grant 7523004 (Algorithmic Data Analysis).

Chapter 5 by V. Berthé, A. Siegel and J. Thuswaldner
Substitutions, Rauzy fractals, and tilings

This chapter focuses on a multiple tiling associated with a primitive substitution σ. We restrict to the case where the inflation factor of the substitution σ is a unit Pisot number. This multiple tiling is composed of tiles which are given by the unique solution of a set equation expressed in terms of a graph associated with the substitution σ: these tiles are attractors of a graph-directed iterated function system (GIFS). Each of these tiles is compact, it is the closure of its interior, it has a non-zero measure and it has a fractal boundary that is also a solution of a graph-directed iterated function system defined by the substitution σ. These tiles are called *central tiles* or *Rauzy fractals*, according to G. Rauzy who introduced them. The aim of this chapter is to list several tiling conditions, relying on the use of various graphs associated with σ.

The authors would like to thank W. Steiner for his efficient help for drawing pictures of fractals, as well as J.-Y. Lee and B. Solomyak for their precious comments on Section 5.7.

Chapter 6 by F. Durand
Combinatorics on Bratteli diagrams and dynamical systems

The aim of this chapter is to show how Bratteli diagrams are used to study topological dynamical systems. Bratteli diagrams are infinite graphs that provide a very efficient encoding of the dynamics that transform some dynamical properties into combinatorial properties on these graphs. We illustrate their wide range of applications through classical notions: invariant measures, entropy, expansivity, representation theorems, strong orbit equivalence, eigenvalues of the Koopman operator.

Chapter 7 by S. Ferenczi and T. Monteil
Infinite words with uniform frequencies, and invariant measures

For infinite words, we study the properties of uniform recurrence, which translates the dynamical property of minimality, and of uniform frequencies, which corresponds to unique ergodicity; more generally, we look at the set of invariant measures of the associated dynamical system. We present some achievements of word combinatorics, initiated by M. Boshernitzan, which allow us to deduce information on these invariant measures from simple combinatorial properties of the words. Then we review some known examples of words with uniform frequencies, and give important examples which do not have uniform frequencies. We finish by hinting how these basic notions have given birth to very deep problems and high achievements in dynamical systems.

The first author wishes to thank the MSRI for its hospitality during the redaction of this chapter. The second author wishes to thank the Poncelet Laboratory and the Asmus Family for their hospitality during the redaction of this chapter.

Chapter 8 by B. Adamczewski and Y. Bugeaud
Transcendence and Diophantine approximation

Finite and infinite words occur naturally in Diophantine approximation when we consider the expansion of a real number in an integer base or its continued fraction expansion. The aim of this chapter is to present several number-theoretical problems that reveal a fruitful interplay between combinatorics on words and Diophantine approximation. For example, if the decimal expansion of a real number viewed as an infinite word on the alphabet $\{0, 1, \ldots, 9\}$ begins with arbitrarily large squares, then this number must be either rational, or transcendental.

Chapter 9 by M. Drmota and P. J. Grabner
Analysis of digital fuctions and applications

The aim of this chapter is to study asymptotic properties of digital functions (like the sum-of-digits function) from different points of view and to survey several techniques that can be applied to problems of this kind. We first focus on properties of average values where we explain periodicity phenomena in the "constant term" or the main term of the corresponding asymptotic expansions. We compare the classical approach by Delange, a Dirichlet series method, and a measure-theoretic method. Secondly, we discuss distributional properties like Erdős-Wintner-type theorems and central limit theorems that work for very general q-additive functions and even if these functions are only considered for polynomial subsequences. These general results are complemented by very precise distributional results for completely q-additive functions which are based on a generating function approach. A final section discusses some further problems like the recent solution of the Gelfond problems on the sum-of-digits function and dynamical aspects of odometers.

The authors are supported by the Austrian Science Foundation FWF, projects S9604 and S9605, parts of the Austrian National Research Network "Analytic Combinatorics and Probabilistic Number Theory".

Chapter 10 by J. Honkala
The equality problem for purely substitutive words

We prove that the equality problem for purely substitutive words is decidable. This problem is also known as the D0L ω-equivalence problem. It was first solved by Culik and Harju. Our presentation follows a simpler approach in which elementary morphisms play an important role. We will also consider the equality problem for sets of integers recognised by finite automata in various ways.

Chapter 11 by V. D. Blondel and R. M. Jungers
Extremal matrix products and the finiteness property

We introduce and study questions related to long products of matrices. In particular, we define the joint spectral radius and the joint spectral subradius which characterise, respectively, the largest and smallest asymptotic rate of growth that can be obtained by forming long products of matrices. Such long products of matrices occur naturally in automata theory due to the possible representation of automata by sets of adjacency matrices.

Joint spectral quantities were initially used in the context of control theory and numerical analysis but have since then found applications in many other areas, including combinatorics and number theory. In the chapter we describe some of their fundamental properties, results on their computational complexity, various approximation algorithms, and three particular applications related to words and languages.