Review of the book
"Combinatorics of Compositions and Words"
by Silvia Heubach and Toufik Mansour
CRC Press, 2010
ISBN:978-1-4200-7267-9

Review by
Serdar Boztaş

June 21, 2011

1 Background on Compositions

The study of compositions of (positive) integers dates to the work of Percy Alexander MacMahon in the late 19th century. As stated by MacMahon "Compositions are merely partitions (of an integer) where the order of occurence of the parts is essential." For example, the partitions of 4 are 4, 31 (representing the equation $3+1=4$), 22, 211 and 1111 while its compositions are $4, 31, 13, 22, 211, 121, 112$ and 1111. Note that the partitions are written in non-ascending order.

Combinatorics on words is a branch of mathematics which applies combinatorics to words and formal languages. The study of combinatorics on words is almost as old as that of compositions, and goes back to the work of Axel Thue on nonrepetitive sequences of symbols at the beginning of the last century.

In particular, the famous Thue-Morse sequence, given below, is an infinite word where no "cubes", i.e., concatenations of nonempty subwords X in the form XXX occur, even though there are infinitely many squares XX within the sequence. Note that we have used spaces to show one (of the many) recurrences that can be used to generate the sequence:

$$0\ 1\ 10\ 1001\ 01101001\ 1001011010010110\ \ldots$$

Note that this sequence is "automatic", i.e., can be generated by a simple automaton. Intriguingly, the real number in $(0,1)$ whose binary expansion is given below

$$0.0110100101101001100101101001011 0 \cdots$$

which is obtained from the Thue-Morse sequence, has been shown to be transcendental.

Research into combinatorics on words and compositions started independently within different branches of mathematics, e.g. number theory, group theory and probability and has applications to combinatorial enumeration, theoretical computer science, automata theory, linguistics, speech processing and bioinformatics. In combinatorics on words one studies words of given length n over an alphabet of size k, usually taken to be $[k] = \{1, 2, \ldots, k\}$, for some $k \geq 2$. Clearly, the two subjects are intimately related and this book brings them together.

One of the main topics covered in the current book is pattern avoidance. In order to make the discussion more concrete, we give an informal definition of a pattern over the alphabet $A = \{1, 2, 3, 4\}$ under the usual ordering: The patterns $12, 13, 14, 23, 24, 34$ all have the reduced form 12, i.e., the first symbol is smaller than the second symbol, while the patterns $123, 124, 134, 234$ all have the reduced form 123 and the patterns $132, 142, 143, 243$ all have the reduced form 132. When one talks of pattern avoidance, it is understood that no subwords (also called substrings in general, and as distinct from subsequences) equivalent to the given reduced pattern occur in the given sequence. For example the pattern 123 does not occur in the sequence 1524354 while the patterns (among others) $132, 213$ and 312 do occur.

2 Summary of the book

We now give a brief overview of the topics covered in the book.

Chapter 1: Introduction starts with a historical perspective and then discusses variations on the main theme, such as compositions with restricted parts, compositions with restricted arrangements, etc. A useful timeline of major research articles in the area is also provided.

Chapter 2: Basic Tools of the Trade embarks on an exposition of techniques for solving recurrence equations, ranging from guessing and checking to the method of Lagrange inversion.

Chapter 3: Compositions provides an exposition of the basic results for compositions, focusing alternately on compositions with restricted positions, and compositions with restricted parts. Subsequently the results are derived for compositions with parts in a general set, and examples of using computer algebra systems for attacking basic problems in compositions are also provided.

Chapter 4: Statistics on Compositions investigates the statistics of rises, falls and levels, as well as focusing on enumeration of patterns and pattern avoidance. The specific case of pattern avoidance for subwords of length 3 is explored in detail. Obtaining results on patterns of length 2 are quite easy, while patterns of length longer than 3 are quite hard in general, if not impossible in some cases.

Chapter 5: Avoidance of Nonsubword Patterns in Compositions considers the same topic as in Chapter 4, but for subsequences, and partially ordered patterns. Some classification results with respect to the Wilf classification of subsequences and generalized patterns are given.

Chapter 6: Words applies the results for compositions to words and also focuses on some specialized techniques for enumeration of words, and pattern avoidance in words.

Chapter 7: Automata and Generating Trees focuses on graph theory basics followed by transfer matrix methods which describe asymptotic behaviour of generating functions.

Chapter 8: Asymptotics for Compositions uses randomized techniques in order to study asymptotics of random compositions and related statistics.

3 What is the book like (style)?

Previous books on this general topic include the three Lothaire (a pseudonym for a group of French mathematicians and computer scientists) volumes [2, 3, 4], and Allouche and Shallit's automatic sequences [1]. The three Lothaire volumes cover most of the same ground as this book, but are written more in the style of a research monograph, while the Heubach-Mansour book is more usable as a textbook, a trait it shares with Allouche and Shallit. However, the relationship between the coverage of the four existing books, and the current book is quite complex in that there are topics which are covered in each one of the books, but not in the others, at least not to the same extent. For example, the current book does not cover stringologic constructions such as suffix trees, tries, etc. which are used in practical implementations of text compression and indexing, since the authors have chosen to focus on topics that are more mathematical in nature. This area is covered in sufficient detail in [3]. On the other hand, the asymptotics of compositions are covered in detail in this book, but not in any of the other four.

4 Would you recommend this book?

In my opinion, this is a book that contains a lot of hidden gems, which need to be explored. It is an advantage that the authors provide fragments of Maple and Mathematica code which would help such explorations. Within one specific context related to cryptography, namely randomness testing, it is always of interest to find an efficient randomness test, and the (sometimes unexpected) peculiarities of distribution that certain subwords can exhibit may help the cryptographer in their search for such tests. Finally, the fact that the authors provided lists of open research problems in each chapter helps to motivate further research into combinatorics on words.

The book is written in an accessible style, and with the help of the five appendices covering Generating function identities; linear algebra; Chebychev polynomials; probability theory and complex analysis, it is quite easy to use for the non-specialist in the area, given a basic computer science and/or mathematical background. It will be a useful reference for the researcher, as well as a very good textbook for a graduate level course in the area. I recommend the book heartily to both the specialists and the beginning researchers in the area.

The reviewer is an academic at RMIT University whose interests cover algorithms, cryptography, sequence design, and coding and information theory.

References

[1] J-P. Allouche and J. Shallit, "Automatic Sequences: Theory, Applications, Generalizations", Cambridge University Press, 2003.

[2] Lothaire (ed.), "Combinatorics on Words", Cambridge University Press, 2nd edition, 1997.

[3] Lothaire (ed.), "Applied Combinatorics on Words (Encyclopedia of Mathematics and its Applications)", Cambridge University Press, 2005.

[4] Lothaire (ed.), "Algebraic Combinatorics on Words (Encyclopedia of Mathematics and its Applications)", Cambridge University Press, 2011.