Ostrowski numeration and repetitions in words

Narad Rampersad

Department of Mathematics and Statistics
University of Winnipeg

- In this talk we give some examples of instances of the Ostrowski numeration system appearing in the study of combinatorics on words.
- Specifically we will look at repetitions in two families of words: balanced words and rich words.

- Here are the basic concepts concerning repetitions.
- Let u be a finite word and write $u = u_0 u_1 \cdots u_{n-1}$, where the u_i are letters.
- A positive integer p is a period of u if $u_i = u_{i+p}$ for all i.
- Write $|u|$ for the length of u: i.e., $|u| = n$.
- Let $e = |u|/p$ and let z be the prefix of u of length p.
- We say that u has exponent e and write $u = z^e$.
- e.g., $01011010 = (01011)^{8/5}$
- A square (resp. cube) is a repetition with exponent 2 (resp. 3)

The critical exponent of an infinite word w is

$$E(w) = \sup\{r \in \mathbb{Q} : \text{there is a finite, non-empty factor of } w \text{ with exponent } r\}.$$

For example, the word

$$w = 012021012102012021\cdots$$

obtained by iterating the substitution

$$0 \to 012, \qquad 1 \to 02, \qquad 2 \to 1$$

contains no squares, but has repetitions with exponents arbitrarily close to 2, so $E(w) = 2$.

Dejean's Theorem

Given an alphabet A of size k, the least critical exponent among all infinite words over A is

$$\begin{cases} 7/4, & k = 3 \\ 7/5, & k = 4 \\ k/(k-1), & k = 2 \text{ or } k \geq 5. \end{cases}$$

What happens if, instead of considering all infinite words, we restrict ourselves to a specific family of infinite words?

- let u be a finite word
- the number of times the letter a appears in u is $|u|_a$
- a word w (finite or infinite) over an alphabet A is balanced if for every $a \in A$ and every pair u, v of factors of w with $|u| = |v|$ we have

$$\big|\, |u|_a - |v|_a \,\big| \leq 1.$$

- the word 0020010201 is not balanced since $|00200|_0 = 4$ and $|10201|_0 = 2$
- the word 01201210210 is balanced

- Balanced words are obtained from Sturmian words.
- Let α be an irrational real number between 0 and 1, called the slope.
- Suppose α has continued fraction expansion $\alpha = [d_0, d_1, d_2, d_3, \ldots]$.

The characteristic Sturmian word with slope α is the infinite word c_α obtained as the limit of the sequence of standard words s_n defined by

$$s_0 = 0, \quad s_1 = 0^{d_1 - 1}1, \quad s_n = s_{n-1}^{d_n} s_{n-2}, \quad n \geq 2.$$

One characteristic Sturmian word is of particular significance. Let $\phi = (1+\sqrt{5})/2$. The Fibonacci word is the characteristic Sturmian word

$$c_\theta = 0100101001001010010100100101001\cdots$$

with slope $\theta := 1/\phi^2 = [0, 2, \overline{1}\,]$. We call the corresponding standard words the finite Fibonacci words:

$$f_0 = 0, \quad f_1 = 01, \quad f_2 = 010, \quad \ldots, \quad f_n = f_{n-1}f_{n-2}$$

- On a binary alphabet, the infinite aperiodic balanced words are exactly the Sturmian words.
- Mignosi and Pirillo (1992) showed that $E(c_\theta) = 2 + \phi$.
- More general results of Damanik and Lenz (2002) and Justin and Pirillo (2001) show that this is minimal over all Sturmian words.
- What about balanced words over larger alphabets?

- An infinite word y has the constant gap property if, for each letter a, there is some number d such that the distance between successive occurrences of a in y is always d.
- This is stronger than being periodic.
- $(0120)^\omega$ is periodic but is not a constant gap word (contains both 00 and 0120)
- $(0102)^\omega$ is a constant gap word

Theorem (Graham 1973; Hubert 2000)

A recurrent aperiodic word x is balanced if and only if x is obtained from a Sturmian word u over $\{0, 1\}$ by:

- replacing the positions containing 0's in u by a periodic sequence y with constant gaps over some alphabet A, and
- replacing the positions containing 1's in u by a periodic sequence y' with constant gaps over some alphabet B, disjoint from A.

e.g., take the Sturmian word

$$u = 0101001010010101001010010101001010010100 \cdots ,$$

$y = (01)^\omega$ and $y' = (2324)^\omega$, then

$$x = 0213012041021302104120130214012031021401 \cdots$$

is balanced.

For each alphabet size k, we wanted to construct balanced words x_k with the least possible critical exponent. For each set of parameters in the table, x_k is constructed from c_α, y, and y' as described above.

k	α	c.f.	y	y'
3	$\sqrt{2}-1$	$[0,\overline{2}\,]$	$(01)^\omega$	2^ω
4	$1/\phi^2$	$[0,2,\overline{1}\,]$	$(01)^\omega$	$(23)^\omega$
5	$\sqrt{2}-1$	$[0,\overline{2}\,]$	$(0102)^\omega$	$(34)^\omega$
6	$(78-2\sqrt{6})/101$	$[0,1,2,1,1,\overline{1,1,1,2}\,]$	0^ω	$(123415321435)^\omega$
7	$(63-\sqrt{10})/107$	$[0,1,1,3,\overline{1,2,1}\,]$	$(01)^\omega$	$(234526432546)^\omega$
8	$(23+\sqrt{2})/31$	$[0,1,3,1,\overline{2}\,]$	$(01)^\omega$	$(23452673254623752643257 6)^\omega$
9	$(23-\sqrt{2})/31$	$[0,1,2,3,\overline{2}\,]$	$(01)^\omega$	$(234567284365274863254768)^\omega$
10	$(109+\sqrt{13})/138$	$[0,1,4,2,\overline{3}\,]$	$(01)^\omega$	$(2345672849632547682943652748 69)^\omega$

- R., Shallit, and Vandomme (2019) showed that:
  - $E(x_3) = 2 + \frac{\sqrt{2}}{2} \approx 2.7071$
  - $E(x_4) = 1 + \frac{\phi}{2} \approx 1.8090.$
- We also conjectured that for $5 \leq k \leq 10$, $E(x_k) = \frac{k-2}{k-3}$.
- We showed that all of these values are minimal, except for x_4, which was done by Peltomäki.
- How does one prove that each word has the claimed critical exponent?

- Here we introduce the Ostrowski α-numeration system.
- Suppose α has continued fraction expansion $\alpha = [0, d_1, d_2, d_3, \ldots]$.
- Let p_n/q_n denote the convergents:

$$\frac{p_n}{q_n} = [0, d_1, d_2, d_3, \ldots, d_n],$$

where

$$p_{-2} = 0, \quad p_{-1} = 1, \quad p_n = d_n p_{n-1} + p_{n-2} \text{ for } n \geq 0;$$
$$q_{-2} = 1, \quad q_{-1} = 0, \quad q_n = d_n q_{n-1} + q_{n-2} \text{ for } n \geq 0.$$

Each non-negative integer N can be represented uniquely as $b_j b_{j-1} \cdots b_0$, where

$$N = \sum_{0 \le i \le j} b_i q_i,$$

where the b_i are digits satisfying:

1. $0 \le b_0 < d_1$,
2. $0 \le b_i \le d_{i+1}$, for $i \ge 1$, and
3. for $i \ge 1$, if $b_i = d_{i+1}$, then $b_{i-1} = 0$.

We call the word $b_j b_{j-1} \cdots b_0$ the canonical Ostrowski α-representation of N.

e.g., if $\alpha = 1/\phi^2 = [0; 2, \overline{1}]$, then the Ostrowski-$\alpha$ numeration system is the classical Zeckendorf numeration system. That is, it is the place-value numeration system where the places have values given by the sequence of Fibonacci numbers:

$$1, 2, 3, 5, 8, \ldots,$$

and canonical representations do not have two consecutive 1's.

e.g., if $\alpha = \sqrt{2} - 1 = [0; \overline{2}]$, then the Ostrowski-$\alpha$ numeration system is the place-value numeration system where the places have values given by the sequence of Pell numbers:

$$1, 2, 5, 12, 29, \ldots,$$

and in any canonical representation, 2's are followed by $0'$'s.

Defining the characteristic Sturmian word of slope α in terms of the Ostrowski numeration system:

Theorem

Let $N \geq 1$ be an integer with Ostrowski α-representation $b_j b_{j-1} \cdots b_0$. Then $c_\alpha[N] = 1$ if and only if $b_j b_{j-1} \cdots b_0$ ends with an odd number of 0's.

- If the continued fraction expansion of α is ultimately periodic (i.e., if α is a quadratic irrational), a finite automaton can easily check if its input is a canonical Ostrowski-α representation.
- The property "ends with an odd number of 0's" can also easily be checked by a finite automaton.
- This means that c_α is an Ostrowski-α-automatic sequence.
- That is, there is a finite automaton with output that outputs $c_\alpha[N]$ when given the Ostrowski-α representation of N as input.

- If c_α is an Ostrowski-α-automatic sequence, it is not hard to show that any word obtained by replacing the 0's and 1's respectively by constant gap sequences y' and y' is also Ostrowski-α automatic.
- The words x_k defined above are Ostrowski-α automatic.
- Given a finite automaton generating an Ostrowski-α automatic sequence, we can use the program Walnut to prove combinatorial properties of the sequence.

- For k-automatic sequences (i.e., sequences generated by a automaton that takes base-k representations as input), Walnut works as follows.
- Given a k-automatic sequence and an expression in first-order logic, where the variables generally represent positions or lengths of factors in the automatic sequence, Walnut will output a new automaton accepting the base-k representations of the natural numbers that satisfy the logical expression.

- Walnut can be extended to work with other numeration systems.
- Important: the addition relation $\{(x, y, z) \in \mathbb{N}^3 : x + y = z\}$ for the numeration system must be recognizable by a finite automaton.
- Clearly this is true for base-k.
- It is also true for the Zeckendorf (Fibonacci) system.

- The word x_4 from the table is Fibonacci-automatic.
- Recall, it is obtained by replacing the 0's (resp. 1's) in the Fibonacci word with $(01)^\omega$ (resp. $(23)^\omega$).

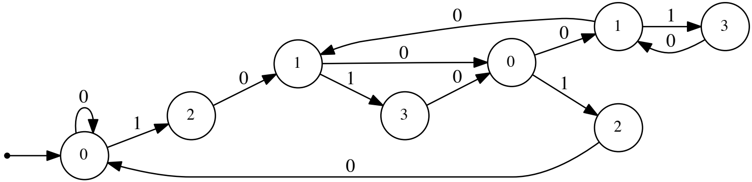

Figure: Fibonacci-base automaton for x_4

Let X denote the automaton for x_4. Using Walnut, we compute the periods p such that a repetition with exponent $\geq 5/3$ and period p occurs in x_4:

```
eval periods_of_high_powers "?msd_fib Ei (p>=1) &
   (Aj (3*j <= 2*p) => X[i+j]=X[i+j+p])";
```

The output of this command is an automaton accepting $0^*1001000^*$; i.e., representations of numbers of the form $F_n + F_{n-3} = 2F_{n-1}$.

```
reg pows msd_fib "0*1001000*";
```

Next we compute pairs (n, p) such that x_4 has a factor of length $n + p$ with period p, and furthermore that factor cannot be extended to a longer factor of length $n + p + 1$ with the same period.

```
def maximal_reps "?msd_fib Ei (Aj (j<n) =>
   X[i+j]=X[i+j+p]) & (X[i+n]!=X[i+n+p])";
```

We now compute pairs (n, p) where p has to be of the form $0^*1001000^*$ and $n + p$ is the longest length of any factor having that period.

```
eval highest_powers "?msd_fib (p >= 1) &
  $pows(p) & $maximal_reps(n,p) &
  (Am $maximal_reps(m,p) => m <= n)";
```

The output of this last command is an automaton accepting pairs (n, p) having the form

$$\binom{0}{0}^* \binom{0}{1} \binom{1}{0} \binom{0}{0} \binom{1}{1} \binom{0}{0} \left\{ \binom{1}{0} \binom{0}{0} \right\}^* \binom{0}{0} \left\{ \epsilon, \binom{1}{0} \right\}.$$

When $p = 2F_{i-1}$ we see that $n = F_i - 2$.

- Maximal repetitions of exponent $\geq 5/3$ in x_4 have exponent of the form $1 + (F_i - 2)/(2F_{i-1})$.
- These exponents converge to $1 + \phi/2$ from below.
- We conclude that x_4 has critical exponent $1 + \phi/2$.

- Recall: the above computations are only possible because addition in the Fibonacci-base is recognizable by a finite automaton.
- What about the other Ostrowski-α numeration systems needed for the other x_i?
- Hieronymi and Terry (2017) showed that addition is indeed recognizable by a finite automaton when α is a quadratic irrational.

- Aseem Baranwal (Master's Thesis 2020) found a simpler way to construct the adder for the Ostrowski numeration with respect to a quadratic irrational α.
- He implemented this in Walnut.

- Baranwal and Shallit (2019) used the adder for the Pell numeration system (Ostrowski-α with $\alpha = \sqrt{2} - 1 = [0; \overline{2}]$) to show that the critical exponent of x_5 is $3/2$.
- Baranwal then extended this to show that for $6 \le k \le 8$, the word x_k as defined in the table above has critical exponent $(k-2)/(k-3)$.
- The computations for $k = 9, 10$ ran out of memory on a computer with 400 GB of RAM.

- Now we switch our attention from balanced words to rich words.
- A palindrome is a word that is equal to its reversal, i.e., it reads the same forwards and backwards.
- A word of length n contains at most n distinct nonempty palindromes.
- Words of length n that contain n distinct nonempty palindromes are called palindrome-rich, or simply rich (or full).
- An infinite word is rich if all of its factors are rich.

- The length 9 word 011101011 is rich.
- It has 9 non-empty palindromes: 0, 1, 11, 111, 01110, 101, 010, 10101, 1101011.
- The length 9 word 011010011 is not rich:
- It only has 8 non-empty palindromes: 0, 1, 11, 0110, 101, 010, 00, 1001

- What is the least possible critical exponent among all infinite rich words on a given alphabet?
- It is necessarily ≥ 2 (Pelantová and Starosta, 2013).
- We resolve this for the binary alphabet.
- Note that all Sturmian words are rich, but the words that achieve the minimal critical exponent among rich words are not Sturmian words.

Let $\Sigma_k = \{0, 1, \ldots, \mathtt{k} - 1\}$. Define $f : \Sigma_3^* \to \Sigma_2^*$ and $g, h : \Sigma_3^* \to \Sigma_3^*$ by

$$
\begin{array}{lll lll lll}
f(0) & = & 0 & f(1) & = & 01 & f(2) & = & 011 \\
g(0) & = & 011 & g(1) & = & 0121 & g(2) & = & 012121 \\
h(0) & = & 01 & h(1) & = & 02 & h(2) & = & 022
\end{array}
$$

- Baranwal and Shallit (2019) showed that $f(h^\omega(0))$ is rich and has critical exponent $2+\sqrt{2}/2$.
- They did so by showing that this word is Pell-automatic: i.e., it is generated by an automaton that takes as input representations of numbers in the Pell numeration system (the Ostrowski-α numeration system corresponding to $\alpha = \sqrt{2}-1 = [0;\overline{2}]$).

- Richness can be verified with Walnut.
- Use the following: A word w is rich if and only if every prefix of w has a unioccurent palindromic suffix.
- The critical exponent computation is done as shown in the previous example.
- Baranwal and Shallit conjectured that this critical exponent of $2 + \sqrt{2}/2$ was minimal over all infinite binary rich words.

Theorem (Currie, Mol, R., 2019)

Let $w \in \Sigma_2^\omega$ be a $14/5$-free rich word. For every $n \geq 1$, a suffix of w has the form $f(h^n(w_n))$ or $f(g(h^n(w_n)))$ for some word $w_n \in \Sigma_3^\omega$.

Theorem (Currie, Mol, R., 2019)

The least critical exponent over all infinite binary rich words is $2 + \sqrt{2}/2$.

- Idea: By the previous theorem, a sequence with the least critical exponent is one of $f(h^\omega(0))$ or $f(g(h^\omega(0)))$.
- Both are rich and have critical exponent $2+\sqrt{2}/2$.
- Baranwal and Shallit, 2019 showed the first, using Walnut;
- Currie, Mol., and R., 2020 showed the second, using other techniques.
- The main idea for the latter was the observation of Pelantová that these words are complementary symmetric Rote words.

References

- For most of this, see Aseem Baranwal's Master's thesis (ask Jeff Shallit for a copy). Also, see:

- N. Rampersad, J. Shallit, E. Vandomme, "Critical exponents of infinite balanced words", Theoret. Comput. Sci. 777 (2019), 454–463.
- J. Currie, L. Mol, N. Rampersad, "The repetition threshold for binary rich words", Discrete Math. Theoret. Comput. Sci 22 (2020).

The End