Abstract Numeration Systems

Narad Rampersad

University of Liège

Sums of three squares

- $7, 15, 23, 28, 31, 39, \ldots$
- These numbers cannot be written as a sum of three squares.
- Is there a pattern?

The three-squares theorem

Theorem (Legendre–Gauss 1798)

A number n is the sum of three squares if and only if n is not of the form $4^a(8m + 7)$.

A language-theoretic version

Theorem (Legendre–Gauss 1798)

A number n is the sum of three squares if and only if the binary representation of n is not of the form $(0+1)^*111(00)^*$.

Recognizable sets

- A set $X \subseteq \mathbb{N}$ is k-recognizable (or k-automatic) if the language $[X]_k$ consisting of the base-k representations of the elements of X is accepted by a finite automaton.

The natural numbers

- The set $\mathbb{N}$ is k-recognizable for all k.
- $[\mathbb{N}]_k$ is the regular language

$$\{1, \ldots, k-1\}\{0, 1, \ldots, k-1\}^* \cup \{\epsilon\}.$$

Examples in base 2

- The Thue–Morse set

$$\{n \in \mathbb{N} : [n]_2 \text{ contains an odd number of 1's}\}$$

  is 2-recognizable.
- The set of powers of 2 is 2-recognizable.

A “gap” theorem

Theorem (Eilenberg)

Let $k \geq 2$ be an integer. A k-recognizable set

$$X = \{x_0 < x_1 < \cdots\}$$

of non-negative integers satisfies either

$$\limsup_{n \to \infty} (x_{n+1} - x_n) < \infty$$

or

$$\limsup_{n \to \infty} \frac{x_{n+1}}{x_n} > 1.$$

The set of squares

- The set $\{n^2 : n \in \mathbb{N}\}$ of squares is not k-recognizable:

$$\limsup_{n\to\infty}((n+1)^2 - n^2) = \infty$$

and

$$\limsup_{n\to\infty} \frac{(n+1)^2}{n^2} = 1.$$

The prime numbers

- The set of prime numbers is not k-recognizable for any k.
- There can be arbitrarily large gaps between sucessive prime numbers.
- If p_n is the n-th prime, the Prime Number Theorem implies that

$$p_{n+1}/p_n \to 1.$$

Changing the base

- Recognizability depends on the base.
- The powers of 2 are not recognizable in base 3.
- Which sets are recognizable in all bases?
- Two numbers k and ℓ are multiplicatively independent if $k^m = \ell^n$ implies $m = n = 0$.
- A set is ultimately periodic if it is a finite union of arithmetic progressions.

Sets recognizable in all bases

Cobham's Theorem 1969

Let $k, \ell \geq 2$ be two multiplicatively independent integers and let $X \subseteq \mathbb{N}$. The set X is both k-recognizable and ℓ-recognizable if and only if it is ultimately periodic.

Determining periodicity

- Given an automaton accepting a k-recognizable set, can we tell if this set is ultimately periodic?
- Honkala (1986) showed that there is an algorithm to solve this problem.
- This result was often reproved: Muchnik (1991), Fagnot (1997), Allouche et al. (2009).
- Leroux (2005) gave a polynomial time algorithm.

The characteristic sequence of the powers of 2

- Iterate the 2-uniform map

$$a \mapsto ab, b \mapsto bc, c \mapsto cc$$

to get the infinite sequence

$$abbcbcccbcccccccbcccccccccccccccbcc \cdots .$$

- Now recode by $a, c \mapsto 0$; $b \mapsto 1$:

$$01101000100000001000000000000000100 \cdots .$$

Automatic sequences

- A sequence is k-automatic if it is generated by first iterating a k-uniform map and then renaming some of the symbols.
- E.g.: the Thue–Morse sequence is generated by the 2-uniform map $0 \mapsto 01;\ 1 \mapsto 10$:

$$0 \to 01 \to 0110 \to 01101001 \to 0110100110010110 \to \cdots .$$

A morphic characterization of k-recognizability

Theorem (Cobham 1972)

Let $k \geq 2$. A set $X \subseteq \mathbb{N}$ is k-recognizable if and only if its characteristic sequence is k-automatic.

A logical characterization of k-recognizability

Theorem (Büchi–Bruyère)

Let $V_k(n)$ be the largest power of k that divides n. A set is k-recognizable if and only if it is definable in the first order theory $\langle \mathbb{N}, +, V_k \rangle$.

Algebraicity of formal power series

Theorem (Christol 1979)

Let p be a prime and let $\mathbf{a} = (a_n)_{n \geq 0}$ be a sequence over $\{0, 1, \ldots, p-1\}$. Then

$$\sum_{n \geq 0} a_n X^n$$

is algebraic over $\mathbb{F}_p(X)$ if and only if $\mathbf{a}$ is a p-automatic sequence.

Positional numeration systems

- A positional numeration system is an increasing sequence of integers $U = (U_n)_{n \geq 0}$ such that
  - $U_0 = 1$ and
  - $C_U := \sup_{n \geq 0} \lceil U_{n+1}/U_n \rceil < \infty$.

The Fibonacci numeration system

- Let $U = (U_n)_{n\geq 0}$ be the sequence of Fibonacci numbers.
- I.e., $U_{n+2} = U_{n+1} + U_n$ and $U_0 = 1$, $U_1 = 2$.
- The greedy representation of 13 is 100000, since

$$13 = 1 \cdot 13 + 0 \cdot 8 + 0 \cdot 5 + 0 \cdot 3 + 0 \cdot 2 + 0 \cdot 1.$$

- 13 also has the non-greedy representation 11000.
- The language of greedy representations is

$$1\{0, 01\}^* \cup \{\epsilon\}.$$

Linear numeration systems

- A set X of integers is U-recognizable if the language $[X]_U$ of greedy representations is accepted by a finite automaton.
- A numeration system $U = (U_n)_{n \geq 0}$ is linear if it satisfies a linear recurrence over $\mathbb{Z}$.
- Introduced and studied by Fraenkel (1985).
- If $\mathbb{N}$ is U-recognizable, then U is linear (Shallit).

A system where $\mathbb{N}$ is not recognizable

- Let U be defined by $U_n = (n+1)^2$ for $n \geq 0$.
- Then U satisfies the linear recurrence

$$U_{n+3} = 3U_{n+2} - 3U_{n+1} + U_n.$$

- Suppose $[\mathbb{N}]_U$ regular. Then

$$[\mathbb{N}]_U \cap 10^*10^* = \{10^a 10^b \in \{0,1\}^* : b^2 < 2a + 4\}$$

would also be regular, which is easily shown to be false.

Bertrand numeration systems

- A numeration system $U = (U_n)_{n \geq 0}$ is a Bertrand numeration system if it has the following property:

  a word w is in $[\mathbb{N}]_U$ if and only if $w0$ is in $[\mathbb{N}]_U$.

Examples of Bertrand systems

- the integer base-k numeration systems
- the Fibonacci numeration system

A non-Bertrand numeration system

- Change the initial conditions of the Fibonacci recurrence $U_{n+2} = U_{n+1} + U_n$ to $U_0 = 1$, $U_1 = 3$:
- the greedy representation of the number 2 is the word 2
- the greedy representation of the number 6 is the word 102, not the word 20.

β-expansions

- Bertrand systems are linked with β-expansions.
- Let $\beta > 1$ be a real number.
- The β-expansion of $x \in [0, 1]$, denoted $d_\beta(x) = (t_i)_{i \geq 1}$, is the lexicographically largest sequence of non-negative integers such that

$$x = \sum_{i=1}^{\infty} t_i \beta^{-i}.$$

Parry numbers

- If

$$d_\beta(1) = t_1 \cdots t_m 0^\omega,$$

with $t_m \neq 0$, then we say that $d_\beta(1)$ is finite.

- In this case we define

$$d^*_\beta(1) = (t_1 \cdots t_{m-1}(t_m - 1))^\omega.$$

- Otherwise, we define $d^*_\beta(1) = d_\beta(1)$.
- If $d^*_\beta(1)$ is ultimately periodic, then β is called a Parry number.

Systems with a dominant root

- Let U be a linear numeration system.
- If
$$\lim_{n\to\infty} \frac{U_{n+1}}{U_n} = \beta$$
for some real $\beta > 1$, then U satisfies the dominant root condition.
- β is the dominant root.

A necessary condition for recognizability

Theorem (Hollander 1998)

Let U be a linear numeration system with dominant root β. If $\mathbb{N}$ is U-recognizable, then β is a Parry number.

- Hollander also gave a much stronger result.

A characterization of the Bertrand systems

Theorem (Bertrand 1989)

Let $U = (U_n)_{n\geq 0}$ be a numeration system. Let Fact(D_β) denote the set of factors occurring in the β-expansions of the real numbers in $[0,1)$. There exists a real number $\beta > 1$ such that $0^*[\mathbb{N}]_U =$ Fact(D_β) if and only if U is a Bertrand numeration system. In this case, if $d^*_\beta(1) = (t_i)_{i\geq 1}$, then

$$U_n = t_1 U_{n-1} + \cdots + t_n U_0 + 1.$$

Obtaining a Bertrand system from a Parry number

- If β is a Parry number, then U is a linear recurrence sequence and β is a root of its characteristic polynomial.
- Every Parry number β has an associated canonical numeration system.
- The language of the canonical numeration system associated with β is $\text{Fact}(D_\beta)$.
- It is a regular language.
- I.e., $\mathbb{N}$ is U-recognizable.

Pisot systems

- A Pisot number is a real algebraic integer greater than one such that all of its algebraic conjugates have absolute value less than one.
- A Pisot system is a linear numeration system whose characteristic polynomial is the minimal polynomial of a Pisot number.

Theorem (Frougny–Solomyak; Bruyère–Hansel)

Let U be a Pisot system. Then $\mathbb{N}$ is U-recognizable.

Determining periodicity

- Given a linear numeration system U and an automaton accepting a U-recognizable set, can we tell if this set is ultimately periodic?
- For Pisot systems, yes. The proof of Allouche et al. (2009) for the integer bases can be applied here (because in Pisot systems, addition is "recognizable").
- The problem is open in general.
- Partial results due to Bell et al. (2009).

A more general class of numeration systems

- Previously, we considered positional systems and then restricted our attention to those that give rise to a regular numeration language.
- Now we consider an arbitrary regular language and use it to define a numeration system (Lecomte and Rigo 2001).

Abstract numeration systems

- An abstract numeration system is a triple $S = (L, \Sigma, <)$:
- $(\Sigma, <)$ is a totally ordered alphabet.
- The numeration language L is an infinite regular language over Σ.
- $[\cdot]_S \colon \mathbb{N} \to L$ is a bijection mapping $n \in \mathbb{N}$ to the $(n+1)$-th word of L in the genealogical order.
- $X \subseteq \mathbb{N}$ is S-recognizable if $[X]_S = \{[n]_S \colon n \in X\}$ is regular.

A general framework

- The base-k system is an abstract numeration system with numeration language

$$\{1,\ldots,k-1\}\{0,1,\ldots,k-1\}^* \cup \{\epsilon\}.$$

- The Fibonacci system is an abstract numeration system with numeration language

$$1\{0,01\}^* \cup \{\epsilon\}.$$

Another abstract numeration system

- Recall: the set $\{n^2 : n \in \mathbb{N}\}$ of squares is not k-recognizable for any k.
- The set of squares is S-recognizable for the abstract numeration system

$$S = (a^*b^* \cup a^*c^*, \{a, b, c\}, a < b < c).$$

- The language of representations of the squares is the regular language a^*.

Polynomial sequences

Theorem (Rigo; Strogalov)

For any polynomial $P \in \mathbb{Q}[x]$ such that $P(\mathbb{N}) \subseteq \mathbb{N}$, there exists S such that P is S-recognizable.

Recgonizability of periodic sets

Theorem (Lecomte and Rigo 2001)

Let S be an abstract numeration system. Every ultimately periodic set is S-recognizable.

The characteristic sequence of the squares

- Recall: the set of squares is recognizable in the system

$$S = (a^*b^* \cup a^*c^*, \{a, b, c\}, a < b < c).$$

- Its characteristic sequence can be generated using the non-uniform morphism

$$h : a \mapsto abcc, b \mapsto bcc, c \mapsto c$$

and a coding

$$g : a, b \mapsto 1, c \mapsto 0.$$

The characteristic sequence of the squares

- We have

$$a \longrightarrow abcc \longrightarrow abccbcccc \longrightarrow abccbcccccc \longrightarrow \cdots$$

  and when we recode we obtain the sequence

$$1100100001000000100000000 \cdots .$$

- A sequence is morphic if it is generated by first iterating a morphism and then renaming some of the symbols.

A morphic characterization of S-recognizability

Theorem (Rigo and Maes 2002)

Let $X \subseteq \mathbb{N}$. Then there exists an abstract numeration system S such that X is S-recognizable if and only if the characteristic sequence of X is morphic.

Determining periodicity

- Given an abstract numeration system S and an automaton accepting a S-recognizable set, can we tell if this set is ultimately periodic?
- The problem is open for the restricted case of linear numeration systems, so it is open here as well.
- In view of the equivalence of S-recognizable sets and morphic sequences, the periodicity question is equivalent to the analogous problem for morphic sequences.
- This is the HD0L periodicity problem, a longstanding open problem in combinatorics on words.

Conclusion

- Abstract numerations generalize the integer base systems and the linear numeration systems.
- They include these systems as special cases.
- In the general case, certain interesting properties are preserved (e.g., recognizability of periodic sets).
- Certain properties are (possibly) lost (e.g., a logical characterization of recognizability).
- A generalization of S-recognizable sets to higher dimensions has also been studied.

The End