DISTRIBUTION OF DIGITS AND OTHER BRANCHES OF MATHEMATICS

MEHDI GOLAFSHAN

Abstract. In this talk, we will discuss the distribution of digits and compute the complexity function and arithmetic complexity. In addition, several open and interesting problems will be raised.

1. Introduction

First of all, we investigate unipotent dymamics on a torus and apply it to the following problem. For an integer k, consider the sequence of digits $(a_n)_{n>0}$, where a_n is the first digit in the decimal representation of 2 to the power n^k. For $k = 1$, we study the sequence

$$1248136125124813612512481 36 \cdots .$$

For $k = 2$, we get

$$121563651212271519342541213 \cdots .$$

In particular, we are interested in the number of factors of length n that may occur in such a sequence (i.e., the subsequences made of n consecutive digits). Digital problems of this type in Number theory are well-known to be difficult, e.g., in the literature, least non-zero digit of $n!$ in base 12 (Deshouillers et al. [1]) or digits of n^n have been investigated. In particular, this permitted me to be familiarized with notions coming from symbolic dynamics [2].

Furthermore, in [3], the leading digits of the Mersenne numbers $M_n = 2^{p_n} - 1$, where p_n is the n^{th} prime were considered.

2. Main results

This work investigates leading digit sequences from a complexity perspective. There are numerous approaches to quantify the *complexity* of a word over a finite alphabet; for instance see [4, 5], and [6] for a summary of various complexity measures. The complexity function will be used as our primary measure of complexity.

2.1. **Number theory and digits (joint work with A. Kanel-Belov).**

Many classical problems of analytical number theory are associated with the study of the sequences of fractional parts of the values of polynomials at integer points. These sequences play an important role in a number of other fields, in particular, in ergodic theory, complexity theory, information transfer theory and [7]. We use the methods of symbolic dynamics to study the fractional parts of the values of the polynomial, namely, we study the unipotent transformation of the torus.

Definition 1. The *factor complexity* or *complexity function* of a finite or infinite word $\mathbf{w}$ is the function $n \mapsto \mathrm{P}_{\mathbf{w}}(n)$, which, for each integer n, give the number $\mathrm{P}_{\mathbf{w}}(n)$ of distinct factors of length n in that word.

This function was first implemented by Hedlund and Morse in 1938.

Definition 2. For every non-zero real number x, *the first significant (decimal) digit* of x, denoted by $D(x)$, is the unique integer $j \in [\![1, 9]\!]$ satisfying

$$10^k j \leqslant |x| < 10^k (j+1)$$

for some (necessarily unique) $k \in \mathbb{Z}$.

Theorem 1. Let $\mathrm{P}(n)$ be a polynomial with an irrational leading coefficient. Let $\mathbf{w}$ be an infinite word where $\mathbf{w}_n$ is the firs significant digits in the binary expansion of $\mathrm{P}(n)$. Then there is a polynomial $\mathrm{Q}(k)$ that depends only on $\deg(\mathrm{P})$, such that $\mathrm{Q}(k) = \mathrm{P}_{\mathbf{w}}(k)$ for all sufficiently large k.

We use this theorem to prove the following results.

2.2. **Ergodic theory and digits (joint work with I. Mitrofanov).**

The following Olympiad problem is widely known, it was designed by Kanel-Belov.

Consider the sequence whose n^{th} term is the most significant digit of 2^n. Show that the number of different words made of 13 consecutive digits is 57.

It can be shown that the number of different words with length n is an affine function in n. Moreover, in [8] it was shown that for all square-free bases $b \geqslant 5$ and all $a \in \mathbb{Q}$, where a is not an integer power of b, the number of different factors of length n in the sequence of the most significant digit of a^k in base-b representation is an affine function.

Theorem 2. For a positive integer d consider the sequence $\mathbf{w}$, where $\mathbf{w}_n$ is the first significant digit of the number 2^{n^d} in decimal representation. Then there is a polynomial $\mathrm{P}(k)$ of degree $d(d+1)/2$ such that $\mathrm{P}(k) = \mathtt{p}_{\mathbf{w}}(k)$ for all sufficiently large k.

The following result is obtained immediately.

Corollary 3. Let a be a fixed positive integer number and not a power of 10. Then Theorem 2 is correct for a^{n^d}.

2.3. **Geometry and digits (joint work with J. Cassaigne).** The arithmetic complexity of an infinite word is the function that counts the number of words of a specific length composed of letters in arithmetic progression (and not only consecutive) In fact, it's a generalization of the complexity function. This concept was introduced by Sergey Avgustinovich and Anna Frid in [9].

Definition 3. Let $\mathbf{w} \in \mathcal{A}^{\mathbb{N}}$ such that $\mathbf{w} = a_0 a_1 \cdots a_n \cdots$, where $a_i \in \mathcal{A}$. We call arithmetic closure of W all

$$A(\mathbf{w}) = \{a_i a_{i+d} a_{i+2d} \cdots a_{i+kd} \mid d \geqslant 1, k \geqslant 0\}.$$

The arithmetic complexity of $\mathbf{w}$ is the function $a_{\mathbf{w}}$ mapping n to the number $a_{\mathbf{w}}(n)$of words in length n in $A(\mathbf{w})$.

Proposition 4. Let $\alpha \in \mathbb{R}_{>0} \setminus \{10^x : x \in \mathbb{Q}\} = \mathfrak{R}$. Then consider $\mathbf{w}_\alpha$ as the word of the most significant digits of α^n.Then

I) $\mathtt{p}_{\mathbf{w}_\alpha} = \theta(n)$ and $a_{\mathbf{w}_\alpha} = \theta(n^3)$.

II) This arithmetic complexity is independent from α. On the other side, $A(\mathbf{w}_\alpha)$ as well is independent from α. Hence, for any $\alpha, \beta \in \mathfrak{R}$, we have $A(\mathbf{w}_\alpha) \subseteq A(\mathbf{w}_\beta)$.

References

[1] J. Deshouilers. "Yet another footnote to the least nonzero digit of $n!$ in base 12." *Unif. distrib. theory.* (2017): 163–167.

[2] A. Kanel-Belov, G. V. Kondakov, and I. V. Mitrofanov. "Inverse problems of symbolic dynamics." *Banach Cent. Publ.* **94**, no. 1 (2011): 43-60.

[3] Zh. Cai, M. Faust, A. J. Hildebrand, J. Li, and Y. Zhang. "Leading digits of Mersenne numbers." *Exp. Mathematics.*

[4] S. Ferenczi. "Complexity of sequences and dynamical systems." *Discrete Math.* **206** (1999): 145–154.

[5] J.-P. Allouche. "Surveying some notions of complexity for finite and infinite sequences." *RIMS Kokyuroku Bessatsu.* **B34** (2012): 27-37.

[6] T. Kamae. "Behavior of various complexity functions." *Theor. Comput. Sci.* **420** (2012): 36–47.

[7] I.M. Vinogradov. "On the question of the distribution of fractional parts of a polynomial." *Izv.* **27**, 6 (1961): 749–754. (Russian)

[8] X. He, A. J. Hildebrand, Y. Li, and Y. Zhang. "Complexity of Leading Digit Sequences." *Discrete Math. Theor. Comput. Sci.* **1**, no. 1 (2020).

[9] S.V. Avgustinovich, D.G. Fon-Der-Flaass, and A.E. Frid. "Arithmetical complexity of infinite words." *Words, Languages and Combinatorics III (Proc. 3rd ICWLC, Kyoto, March 2000), World Scientific, Singapore.* (2003) :51–62.

Mehdi Golafshan: Department of Mathematics, University of Liège, Liège, Belgium

Email address: `mgolafshan@uliege.be`