AUTOMATIC SEQUENCES, GENERALISED POLYNOMIALS, AND NILMANIFOLDS

JAKUB BYSZEWSKI (JOINT WORK WITH JAKUB KONIECZNY)

Description of results

Automatic sequences are sequences whose n-th term is produced by a finite state machine from base k digits of n. By definition, automatic sequences can take only finitely many values. Allouche and Shallit [AS], [AS2] have generalized the notion of automatic sequences to a wider class of regular sequences and demonstrated their ubiquity and links with multiple branches of mathematics and computer science. The problem of demonstrating that a certain sequence is or is not automatic or regular has been widely studied, particularly for sequences of arithmetic origin. We continue this study for sequences that arise from generalized polynomials, i.e. expressions involving algebraic operations and the floor function via dynamical and ergodic methods. This is possible because by the work of Bergelson and Leibman generalized polynomials are strongly related to dynamics on nilmanifolds. The results obtained lead to a number of interesting questions concerning zero sets of generalized polynomials that we hope will be of independent interest.

In [AS2, Theorem 6.2] it is proved that the sequence $(f(n))_{n\geq 0}$ given by $f(n) = \lfloor \alpha n+\beta \rfloor$ for real numbers α, β is regular if and only if α is rational. The method used there does not immediately generalise to higher degree polynomials in n, but the proof implicitly uses rotation on a circle by an angle of $2\pi\alpha$. Replacing the rotation on a circle by a skew product transformation on a torus (as in Furstenberg's proof of Weyl's equidistribution theorem), we easily obtain the following result.

Theorem A. *Let $p \in \mathbb{R}[x]$ be a polynomial. Then the sequence $f(n) = \lfloor p(n) \rfloor, n \geq 0$ is regular if and only if all the coefficients of p except possibly for the constant term are rational.*

In fact, we show the stronger property that for an integer $m \geq 2$ the sequence $f(n) \bmod m$ is not automatic unless all the coefficients of p except for the constant term are rational, in which case it is periodic. It is natural to inquire whether a similar result can be proven for more complicated expressions involving the floor function such as e.g. $f(n) = \lfloor \alpha \lfloor \beta n^2 + \gamma \rfloor^2 + \delta n + \varepsilon \rfloor$. Such sequences are called generalized polynomial and have been intensely studied. The main motivation for this project is the following conjecture.

Conjecture A. *Suppose that a sequence f is simultaneously automatic and generalised polynomial. Then f is ultimately periodic.*

(We say that a sequence f is *ultimately periodic* if it coincides with a periodic sequence except at a finite set.)

We are able to partially resolve this conjecture. First of all, we prove that the conjecture holds except on a set of density zero. In fact, in order to obtain such a

result, we only need a specific property of automatic sequences. For the purpose of stating the next theorem, let us say that a sequence $f\colon \mathbb{N} \to X$ is *weakly periodic* if for any restriction of f to an arithmetic sequence, $f'(n) = f(an+b)$, $a \in \mathbb{N}$, $b \in \mathbb{N}_0$, there exist $q \in \mathbb{N}$, $r, s \in \mathbb{N}_0$ with $r \neq s$, such that $f'(qn + r) = f'(qn + s)$. Of course, any periodic sequence is weakly periodic, but not conversely. All automatic sequences are weakly periodic, which follows from the finiteness of kernels.

Theorem B. *Suppose that a sequence $f\colon \mathbb{N} \to \mathbb{R}$ is weakly periodic and generalised polynomial. Then there exists a periodic function p and a set $Z \subset \mathbb{N}$ of (upper Banach) density zero such that $f(n) = p(n)$ for $n \in \mathbb{N} \setminus Z$.*

To obtain stronger bounds on the size of the exceptional set Z, we need to restrict to automatic sequences and exploit some of their finer properties.

Theorem C. *Suppose that a sequence $f\colon \mathbb{N} \to \mathbb{R}$ is automatic and generalised polynomial. Then there exists a periodic function p and a set $Z \subset \mathbb{N}$ such that $f(n) = p(n)$ for $n \in \mathbb{N} \setminus Z$ and*

$$\sup_M |Z \cap [M, M + N)| = O\left(\log^C N\right)$$

as $N \to \infty$ for a certain constant C.

While Theorem C does not resolve Conjecture A, our proof thereof greatly restricts the number of possible counterexamples. In fact, in order to prove Conjecture A, it would suffice to prove that the characteristic sequence of powers of an integer $k \geq 2$ given by

$$g_k(n) = \begin{cases} 1, & \text{if } n = k^t \text{ for some } t \geq 0; \\ 0, & \text{otherwise} \end{cases}$$

is not generalized polynomial.

Theorem D. *Let $k \geq 2$ be an integer. Then one of the following statements holds:*

(i) All sequences which are simultaneously k-automatic and generalised polynomial are ultimately periodic.

(ii) The characteristic sequence g_k of the powers of k is generalised polynomial.

References

[AS] Jean-Paul Allouche and Jeffrey Shallit, *The ring of k-regular sequences*, Theoret. Comput. Sci., 98(2):163–197, 1992.

[AS2] Jean-Paul Allouche and Jeffrey Shallit, *The ring of k-regular sequences*, II. Theoret. Comput. Sci., 307(1):3–29, 2003.

Efficient repetition-free strings generator

Anton Chaplygin

Ural Federal University

Repetitions in strings (words) is a popular topic in both combinatorics of words and stringology. A novel approach in this area is the generation of repetition-free strings by local resampling algorithms, inspired by Moser and Tardos's constructive proof of the Lovasz local lemma [2].

Let us be more precise. Consider a string $s = (uv)^k u$ for some integer $k \geq 1$, we say s is a repetition with a period $p = |uv|$ and an exponent $\beta = \frac{|s|}{p}$, or, simply, β-repetition. One may notice that β can be any rational and $\beta \geq 1$. A string is β-repetition-free if it does not contain a repetition with exponent larger or equal to β.

A well known example of β-repetitions is squares ($\beta = 2$). A generator of square-free strings from a random source have been proposed by Arseny Shur in [3] along with the asymptotic formula giving the expected number of random letters used by the generator to construct a square-free string of length n depending on a fixed alphabet size.

In my work, I extended the mentioned generator to produce β-repetition-free strings for an arbitrary fixed exponent $\beta > 1$ and fixed alphabet Σ. The generation algorithm uses the repetition detector by Dmitry Kosolobov [1] which finds the earliest occurrence of a β-repetition in the string in the online fashion. It successively appends random symbols to the end of the string and checks for a β-repetition occurrence. As soon as an occurrence is detected, some its suffix is removed and the structure is rolled back to the corresponding state. This algorithm works in $O(N \cdot \log n)$ time, where n is a length of generated string and N is a number of randomly generated symbols.

I implemented the algorithm as a software and used it to study the structure of infinite β-repetition-free languages for different values of β and different alphabetic sizes. The most important characteristic of a language, estimated by this software, is the conversion coefficient $\lim_{n\to+\infty} E(\frac{N}{n})$, where N is number of random symbols required by the algorithm to

generate a β-repetition-free string of length n. The larger this coefficient is, the harder it is for the random process to choose an infinite branch in the tree of all β-repetition-free words, avoiding dead ends.

The program runs the algorithm for different values of n up to 10^6, multiple times for each value, taking the mean values to estimate the expectation. The main experimental results are as follows:

- For $\beta = 2$ the coefficients agree with the results of [3]
- For $\beta = (\frac{|\Sigma|}{|\Sigma|-1})^+$ with $|\Sigma| \geq 5$ (minimal infinite repetition-free languages; "+" means that the exponent exactly β is permitted) the coefficients seem to tend to a limit ≈ 3.5 as $|\Sigma| \to \infty$
- $\Sigma = 4$, $\beta = (\frac{7}{5})^+$ (minimal quaternary repetition-free language) the biggest coefficient was observed ≈ 300
- $\Sigma = 3$, $\beta = (\frac{7}{4})^+$ (minimal ternary repetition-free language) is the most intriguing, because this is the only language for which the algorithm fails to produce a string longer than several hundred symbols. The behaviour of the algorithm on this language will be our next object of study.

References

[1] D. Kosolobov. Online detection of repetitions with backtracking. In *CPM 2015*, volume 9133 of *LNCS*, pages 295–306. Springer International Publishing, 2015.

[2] R. Moser and G. Tardos. A constructive proof of the general lovász local lemma. *J. of ACM*, 57, January 2010.

[3] A. M. Shur. Generating square-free words efficiently. *Theoretical Computer Science*, 601:67–72, 2015.

The Černý conjecture and 1-contracting automata

Henk Don

1 Introduction

Let $\mathscr{A} = (Q, \Sigma, \delta)$ be a deterministic finite automaton (DFA), where Q denotes the state set, Σ the input alphabet, and $\delta : Q \times \Sigma \to Q$ the transition function. We denote the set of finite words over Σ by $\Sigma^\star$. The transition function δ extends uniquely to a function $\delta : Q \times \Sigma^\star \to Q$.

The automaton $\mathscr{A}$ is called *synchronizing* if there exists a word $w \in \Sigma^\star$ and $q \in Q$ such that $\delta(q', w) = q$ for all $q' \in Q$. The word w is then said to be a *synchronizing word* for $\mathscr{A}$.

The following longstanding conjecture is due to Černý ([1], 1964):

Conjecture 1. If $\mathscr{A}$ is a synchronizing n-state automaton, then there exists a synchronizing word for $\mathscr{A}$ of length at most $(n-1)^2$.

2 1-contracting automata

In this talk we look at n-state automata in which every $(n-1)$-subset of the state set Q is reachable from Q. Such automata will be called 1-contracting. A word with the property that it maps Q to an $(n-1)$-subset of Q is called 1-deficient. If w is a 1-deficient word, the state that is not in the image of w is said to be the *excluded* state. There also must be a unique state in the image which is reached twice by w. This state will be called the *contracting* state for w.

In a 1-contracting automaton, for every state q there exists a 1-deficient word that excludes q. A collection W of words is called 1-contracting if for all q it contains exactly one word w_q which excludes q. To such a collection we can associate a function σ_W on Q that maps each state q to the unique contracting state for w_q. This function will be called the state map induced by W. If for some 1-contracting collection W the state map is a cyclic permutation on Q, then the automaton is called aperiodically 1-contracting. A formal definition is given below.

Definition 2. Let $\mathscr{A} = (Q, \Sigma, \delta)$ be a DFA with n states. $\mathscr{A}$ is called aperiodically 1-contracting if there exist words $w_1, \ldots, w_n \in \Sigma^*$ and a cyclic order $q_1 \prec q_2 \prec \ldots \prec q_n \prec q_1$ on Q such that for all $i = 1, \ldots, n$ (and interpreting q_{n+1} as q_1)

$$\delta(Q, w_i) = Q \setminus \{q_i\} \quad \text{and} \quad |\delta^{-1}(q_{i+1}, w_i)| = 2.$$

3 Main results

Our main result (which appears in [2]) is the following:

Theorem 3. *Let $\mathscr{A} = (Q, \Sigma, \delta)$ be an aperiodically 1-contracting DFA with n states. If there exists an efficient 1-contracting collection $W \subseteq \Sigma^\star$ for which σ_W is a cyclic permutation on Q, then*

1. *The shortest synchronizing word of $\mathscr{A}$ has length at most $(n-1)^2$.*
2. *For every nonempty set $S \subseteq Q$ of size k, there exists a word w_S of length at most $n(n-k)$ such that $\delta(Q, w_S) = S$.*

So, under the conditions of this theorem, the Černý conjecture holds true. The second statement of the theorem is in fact even stronger, claiming reachability of all subsets of Q, with a quadratic upper bound on the word length.

4 Outline of the talk

In the talk I will introduce Černý's conjecture and demonstrate that it is quite straightforward to find a cubic upper bound for the length of the shortest synchronizing word. Then I will discuss the notion of 1-contracting automata and try to explain the ideas behind this definition. Next, I will present the main result and sketch the proof. If time permits, I will discuss some examples that fit into the class of aperiodically 1-contracting automata.

References

[1] J. Černý. Poznámka k homogénnym experimentom s konečnými automatmi. *Matematicko-fyzikálny časopis, Slovensk. Akad. Vied,* Vol. 14, No. 3, 208–216, 1964.

[2] H. Don. The Černý conjecture and 1-contracting automata. *Electronic Journal of Combinatorics* 23 (3), 2016.

On the Gap Between Separating Words and Separating Their Reversals

Farzam Ebrahimnejad*

Department of Computer Engineering,
Sharif University of Technology

Abstract

The function $\text{sep}(w, x)$ is defined as the size of the smallest deterministic finite automaton that accepts w and rejects x. In 1986, Goralcik and Koubek [2] introduced the separating words problem, which asks for good upper and lower bounds on

$$S(n) \coloneqq \max_{w \neq x \wedge |w|,|x| \leq n} \text{sep}(w, x).$$

Goralcik and Koubek [2] proved $S(n) = o(n)$. Besides, the best known upper bound so far is $O(n^{2/5} (\log n)^{3/5})$, which was obtained by Robson [3]. A recent paper by Demaine et al. [1] surveys the latest results about this problem, and while proving several new theorems, it also introduces three new open problems, all of which have remained unsolved until now. In this paper, we solve the first open problem stated in that paper, which asks whether

$$\left|\text{sep}(w, x) - \text{sep}(w^R, x^R)\right|$$

is bounded or not. We prove that this difference is actually unbounded. In order to do so, for all positive integers $k \in \mathbb{N}$, we will construct two words

$$w = u0^n v, x = u0^{n+(2n+1)!} v,$$

for some $u, v \in \{01, 11\}^+ \left(0^+ \{01, 11\}^+\right)^*$, such that $\text{sep}(w, x) - \text{sep}(w^R, x^R)$ approaches infinity as k approaches infinity. We show that under certain conditions, we can set u, v so that it requires relatively few states to separate w^R, x^R. But while preserving these conditions, we set u, v so that it will require exponentially more states, with respect to k, to separate w and x. We do that by using the regular language G_k, which is described in the paper, and has some interesting characteristics. We show that for all $k \in \mathbb{N}$, there exists $z_k \in G_k$

*Email address: febrahimnejad@ce.sharif.edu

such that if a DFA with less than 2^k states accepts z_k, then it should also accept a word in $\{1,2\}^* - G_k$.

Keywords: Words separation; Finite automata.

This paper has been submitted to the journal of *Theoretical Computer Science*. The full version is available at *arXiv:1605.04835 [cs.FL]*.

References

[1] Erik D. Demaine, Sarah Eisenstat, Jeffrey Shallit, and David A. Wilson. Remarks on separating words. In *Descriptional Complexity of Formal Systems - 13th International Workshop, DCFS 2011, Gießen/Limburg, Germany, July 25-27, 2011. Proceedings*, pages 147–157, 2011.

[2] Pavel Goralcik and Václav Koubek. On discerning words by automata. In *Automata, Languages and Programming, 13th International Colloquium, ICALP86, Rennes, France, July 15-19, 1986, Proceedings*, pages 116–122, 1986.

[3] John M. Robson. Separating strings with small automata. *Inf. Process. Lett.*, 30(4):209–214, 1989.

Densities of sets defined by sum-of-digits function

We present two works, joint with respectively Alexander Prikhodko and Pascal Hubert, where we study properties of densities of sets defined by the sum-of-digits functions in base 2. To be more precise, we are interested, for any $a \in \mathbb{N}$ and any $d \in \mathbb{Z}$, in the following set:

$$E_{a,d} := \{n \in \mathbb{N} \mid s_2(n+a) - s_2(n) = d\}$$

where

$$\begin{array}{rccc} s_2 : & \mathbb{N} & \to & \mathbb{N} \\ & n & \mapsto & \displaystyle\sum_{k=0}^{m} n_k. \end{array}$$

if $n = \displaystyle\sum_{k=0}^{m} n_k 2^k$.

Bésineau proved that such sets admit a partition into arithmetic pogressions and thus that they admit asymptotic densities. The quantity $s_2(n+a) - s_2(n)$ appears naturally when studying the autocorrelation funcrion of some arithmetic functions. Denote $\mu_a(d)$ the asymptotic density of $E_{a,d}$, that is to say:

$$\mu_a(d) = \lim_{N \to +\infty} \frac{1}{N} \left(\# E_{a,d} \cap \{0, ..., N-1\} \right).$$

Remark that for any a, μ_a is a probability measure on $\mathbb{Z}$. We study the asymptotic properties of the probability measures as a goes to $+\infty$. We prove the following property:

Proposition 1. *For any $a \in \mathbb{N}$ and any $d \in \mathbb{Z}$, there exists a finite set $S_{a,d} \subset \{0,1\}^*$, possibly empty, such that:*

$$n \in E_{a,d} \Leftrightarrow \exists w \in \{0,1\}^*, \exists s \in S_{a,d},\ \underline{n} = ws.$$

We give explicite construction of the sets $S_{a,d}$ by recurrence properties on a. This has a direct corollary on μ_a:

$$\mu_{2a}(d) = \mu_a(d),\ \mu_{2a+1}(d) = \frac{1}{2} \left(\mu_a(d-1) + \mu_{a+1}(d+1) \right).$$

This allows not only to compute the probability measures explicitly in a simple manner, but also, by applying Fourier transform on both sides of the equations, we have recurrence relations between the characteristic functions of the probability measures μ_a. This allows us to write the characteristic function of μ_a as a product of matrices. With such a writing we prove the following:

Theorem 2. *Let $l(a)$ denote the number of occurrences of the word "01" in $\underline{a}$. There exists a constant $C > 0$ such that:*

$$\forall a \in \mathbb{N},\ \|\mu_a\|_{l^2(\mathbb{Z})} \leq C\left(l(a)\right)^{\frac{-1}{4}}.$$

Moreover, the study of the characteristic function of μ_a, still as a product of matrices, allows a to prove the following:

Theorem 3. *For any a, the mean of μ_a is 0 and its variance, denoted $Var(\mu_a)$ satisfies:*

$$l(a) - 1 \leq Var(\mu_a) \leq 4l(a) + 2$$

This raised the question as to know whether or not, for a given sequence a_n, the sequence $(\mathrm{Var}(\mu_{a_n})/l(a_n))_{n\in\mathbb{N}}$ converges. We partially answer this question in a joint work with Pascal Hubert.

Let $X = (X_k)_{k\in\mathbb{N}} \in \{0,1\}^{\mathbb{N}}$. Define the associated sequence of integers $(a_X(n))_{n\in\mathbb{N}}$ by:

$$\forall n \in \mathbb{N}, a_X(n) = \sum_{k=0}^{n-1} X_k 2^k.$$

Let $\mathbb{P}$ denote the balanced Bernoulli measure on $\{0,1\}^{\mathbb{N}}$. We have the following:

Proposition 4. *There exists a set $U \subset \{0,1\}^{\mathbb{N}}$ such that $\mathbb{P} = 1$ and:*

$$\forall X \in U,\ \ Var(\mu_{a_X(n)}) \sim \frac{n}{2}.$$

This motivates the following renormalisation. For any $X \in U$ and $n \in \mathbb{N}$, define $\tilde{\mu}_n^X \in l^1\left(\sqrt{\frac{2}{n}}\mathbb{Z}\right)$ by:

$$\forall x \in \sqrt{\frac{2}{n}}\mathbb{Z},\ \tilde{\mu}_n^X(x) = \mu_{a_X(n)}\left(\sqrt{\frac{n}{2}}x\right).$$

We have the following:

Theorem 5. *For any $X \in U$,*

$$\tilde{\mu}_n^X \rightharpoonup \mathcal{N}(0,1).$$

This is proved by computing the moments of $\tilde{\mu}_n^X$ thanks to its characteristic function which is given by a product of matrices and by showing the convergence of these moments towards those of the normal law.

Entropy of topologically mixing subshifts

The entropy is a topological invariant of dynamical systems measuring their complexity. We study the influence of topological mixing hypothesis on the entropy of effective one dimensional subshifts, and two dimensional subshifts of finite type. We would like to present some results about one dimensional subshifts. The main result is that the entropies of $O(n)$-topologically mixing effective one dimensional subshifts are exactly the Π_1-computable numbers. Under some low mixing condition, that is to say $O(f(n))$-mixing with $f(n)$ growing sufficiently slowly, the entropies of such subshifts are all computable numbers. We ask the following questions : can we reduce the gap between these two behaviors ? Could we realize every computable number as the entropy of a low mixing subshift ? We would like also to talk about two dimensional SFTs. We know that $O(log(n))$-mixing SFTs have a computable entropy and have a dense set of periodic points. What happen if we take a greater intensity of mixing ? Can we produce some $O(n)$-mixing aperiodic SFT ? What are the entropies of $O(n)$-mixing SFTs ?

On the number of synchronizing colorings of digraphs

Vladimir V. Gusev

joint work with Elena V. Pribavkina and Marek Szykua

Let $\mathscr{A} = (Q, \Sigma, \delta)$ be a finite deterministic complete automaton with an alphabet Σ, a set of states Q and a transition function δ. The automaton $\mathscr{A}$ is *synchronizing* if there exist a word u and a state p such that for every state $q \in Q$ we have $q \cdot u = p$, where $q \cdot u$ denotes the image of q under the action of u. Any such word u is called *synchronizing* (or *reset*) word for $\mathscr{A}$. The length of the shortest synchronizing word $\mathrm{rt}(\mathscr{A})$ is called *the reset threshold* of $\mathscr{A}$. Synchronizing automata naturally appear in algebra, coding theory, industrial automation, discrete dynamical systems, etc. A brief survey of the theory of synchronizing automata may be found in [8].

Two fundamental problems about synchronizing automata that were intensively investigated in the last decades are the Černý conjecture and the road coloring problem. The former states that the reset threshold of an n-state automaton is at most $(n-1)^2$ [3]. Despite intensive research efforts it remains open for already half a century. The latter problem states a certain connection between primitive digraphs and synchronizing automata, which we will explain shortly, and was recently resolved by Trakhtman [7] after crucial insight by Culik, Karhumäki, and Kari [4]. My talk is devoted to the generalizations of the road coloring theorem.

The road coloring theorem. *The underlying digraph* $\mathcal{G}(\mathscr{A})$ of an automaton $\mathscr{A}$ is a digraph with Q as a set of vertices, and for each $u \in Q, x \in \Sigma$ there is an edge $(u, u \cdot x)$. We allow loops and multiple edges, thus $\mathcal{G}(\mathscr{A})$ has a fixed out-degree equal to the cardinality of the alphabet Σ, i.e., $\mathcal{G}(\mathscr{A})$ is a $|\Sigma|$-out-regular digraph.

Vice versa, given a digraph G with a fixed out-degree k and a finite alphabet Σ with k letters, we can obtain a deterministic finite automaton by distributing the letters of Σ over the edges of G. Any automaton obtained in this way is called a *coloring* of G. A digraph is *primitive* if there exists a number t such that for any two vertices u and v there exists a path from u to v of length exactly t. An automaton is *strongly connected* if its underlying digraph is strongly connected.

Theorem 1 (Road coloring theorem) *A strongly connected digraph G with a fixed out-degree k has a synchronizing coloring if and only if it is primitive.*

This theorem was stated as a conjecture in 1977 [1]. The authors' original motivation comes from symbolic dynamics. Namely, synchronizing coloring defines a morphism from a shift of finite type given by G to a full shift over Σ with special properties, see [2].

The origin of the terminology is as follows. A digraph G represents a network of one-way roads. A coloring of G defines labels of the roads that can be perceived by drivers. If the coloring is synchronizing then the drivers who are unaware of their current location have the following strategy to relocate themselves: they can simply follow roads labelled by a synchronizing word and their final position will be well defined.

Although the road coloring theorem gives an answer for a principal connection between digraphs and synchronizing automata, there are still basic quantitative questions that remain unanswered. Namely, how many synchronizing colorings a primitive digraph G can have and what is the number of synchronizing colorings of an average (or random) digraph? These questions were addressed in our recent works [5, 6].

In my talk I will present two conjectures that generalize the road coloring theorem. Furthermore, I will describe our recent work to prove these conjectures based on the spectral properties of the adjacency matrix $\mathcal{A}(G)$ of a digraph G. Namely, we used the structure of the dominant eigenvector $\vec{v}$ of $\mathcal{A}(G)$ to obtain bounds on the number of synchronizing colorings of G. Using this technique we were able to prove one of the conjectures in a special class of digraphs and reformulate the other.

References

[1] R. L. Adler, L. W. Goodwyn, and B. Weiss. Equivalence of topological Markov shifts. *Israel Journal of Mathematics*, 27(1):49–63, 1977.

[2] M.-P. Béal and D. Perrin. *Handbook of Formal Languages: Volume 2. Linear Modeling: Background and Application*, chapter Symbolic Dynamics and Finite Automata, pages 463–506. Springer Berlin Heidelberg, 1997.

[3] J. Černý. Poznámka k homogénnym experimentom s konečnými automatmi. *Matematicko-fyzikálny Časopis Slovenskej Akadémie Vied*, 14(3):208–216, 1964. In Slovak.

[4] K. Culik, J. Karhumäki, and J. Kari. A note on synchronized automata and road coloring problem. In *Developments in Language Theory*, volume 2295 of *LNCS*, pages 175–185. Springer, 2002.

[5] V. V. Gusev and M. Szykua. On the number of synchronizing colorings of digraphs. In *Implementation and Application of Automata*, volume 9223 of *LNCS*, pages 127–139. Springer, 2015.

[6] Vladimir V. Gusev and Elena V. Pribavkina. On synchronizing colorings and the eigenvectors of digraphs. In Piotr Faliszewski, Anca Muscholl, and Rolf Niedermeier, editors, *41st International Symposium on Mathematical Foundations of Computer Science, MFCS 2016, August 22-26, 2016 - Kraków, Poland*, volume 58 of *LIPIcs*, pages 48:1–48:14. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2016.

[7] A. N. Trahtman. The Road Coloring Problem. *Israel Journal of Mathematics*, 172(1):51–60, 2009.

[8] M. V. Volkov. Synchronizing automata and the Černý conjecture. In *Language and Automata Theory and Applications*, volume 5196 of *LNCS*, pages 11–27. Springer, 2008.

On the numbers of ergodic lifts over ergodic measures for finite-to-one factor maps between shifts of finite type

Uijin Jung

Let f be a finite-to-one factor map between two mixing shifts of finite type X and Y with the same topological entropy. It is well known that there is a natural number d, called the degree of f, such that almost all points in Y have d preimages in X. The degree of f plays an important role in the study of finite-to-one factor maps between symbolic dynamical systems. Since f naturally induces a factor map between the sets of invariant measures of X and Y, it is natural to ask whether there is a relation between the degree and the number of ergodic lifts over a measure on Y. The degree is also an upper bound on the number of ergodic lifts over a fully supported measure on Y. Also, for each fully supported Markov measure on Y, it is known that there is only one invariant measure of X over it. We show that there is also a fully supported ergodic invariant measure on Y for which there exists exactly d ergodic invariant measures mapping to it. Then we discuss on the possible set of the numbers of ergodic lifts over ergodic measures on Y.
This is a joint work with Jisang Yoo.

Quest for Short Identities in Transformation Semigroups and Symmetric Groups

This research was mainly motivated by the separating words problem, formulated by Goralcik and Koubek in 1986 [1], which is stated as follows. Let $Sep(n)$ be the minimum number such that for any two words with length less or equal than n there is a deterministic finite automaton with $Sep(n)$ states that accepts exactly one of them. The problem is to find the asymptotics of the function Sep. The known lower bound for $Sep(n)$ is $\log n + o(\log n)$, and the best upper bound, obtained by Robson [2] is $O(n^{2/5}(\log n)^{3/5})$. There is a version of the problem, in which all automata are permutational. We denote the analog of the function Sep for permutational automata by $Sepp$. In this case the lower bound is the same as the lower bound for Sep, and the upper bound also belongs to Robson [3] and is $O(n^{1/2})$. Such a huge gap suggests that any of these bounds can be very loose.

To improve the lower bound, one needs to find short identities in full transformation semigroups T_k and symmetric groups S_k. This connection stems from the following simple facts:

Fact 1. $Sep(n) \leq k$ iff each identity $u = v$ in T_k satisfies $\max\{|u|, |v|\} > n$.

Fact 2. $Sepp(n) \leq k$ iff each identity $u = v$ in S_k satisfies $\max\{|u|, |v|\} > n$.

Trivially, T_k and S_k satisfy the unary identities

$$x^{k-1+lcm\{1,\ldots,k\}} = x^{k-1} \tag{1}$$

and

$$x^{lcm\{1,\ldots,k\}} = 1 \tag{2}$$

respectively. In addition, it is known that for any non-unary identity there is a binary identity of the same length. Thus, I searched for shorter binary identities for small values of k. The main results, obtained through various computational experiments, are as follows.

For $k = 4$:

- T_4 has no identities shorter than the unary identity (1) (the same result for T_3 was known, as well as the shortest identities for S_3 and S_4).

For $k = 5$:

- T_5 has no identities of length ≤ 40 (this result uses an exhaustive search of identities in S_5 up to length 33 made by K. Startsev; his search reveals just two identities of length 32);

- S_5 satisfies the following identity of length 34:

$$(xy)^{12}(yx)^5 = (yx)^5(xy)^{12} \tag{3}$$

- T_5 satisfies the following identity of length 48, obtained from (3) by adding appropriate prefixes and suffixes to both sides:

$$(xy)^{15}(yx)^5(xy)^4 = (xy)^3(yx)^5(xy)^{16} \tag{4}$$

  Moreover, our guess, supported by some partial search, is that (4) is the shortest identity in T_5.

For $k \geq 6$:

- no short identity in T_k was found yet;

- the shortest identities in S_k of the form

$$(xy)^a(yx)^b = (yx)^b(xy)^a \tag{5}$$

  were found up to $k = 23$; for example, for $k = 23$ such an identity has length 2332920, while $lcm\{1, .., 23\} = 5354228880$;

- even shorter identities in S_k of the form

$$(xy)^a(yx)^b(xy)^c(yx)^d = (yx)^d(xy)^c(yx)^b(xy)^a \tag{6}$$

  were found for $k = 6, \ldots, 12$; for example, S_6 has a unique shortest identity of length 32, and S_7 has an identity of length 76.

These results are a part of the paper [4].

References

[1] P. Goralcik and V. Koubek. On discerning words by automata. In Automata, Languages and Programming, 13th International Colloquium, ICALP86. Proceedings, volume 226 of Lecture Notes in Computer Science, pages 116-122. Springer, 1986.

[2] J. M. Robson. Separating strings with small automata. Inf. Process. Lett., 30(4):209-214, 1989.

[3] J. M. Robson. Separating words with machines and groups. RAIRO Inform. Theor. Appl., 30(1):81-86, 1996.

[4] Andrei A. Bulatov, Olga Karpova, Arseny M. Shur, Konstantin Startsev. Lower Bounds on Words Separation: Are There Short Identities in Transformation Semigroups? arXiv:1609.03199 [math.CO], 2016.

INVARIANT MEASURES OF B-FREE SHIFTS

JAKUB KONIECZNY, MICHAL KUPSA, AND DOMINIK KWIETNIAK

A set $A \subset \mathbb{Z}$ is *periodic* if it is a finite union of infinite arithmetic progressions. Note that characteristic functions of periodic sets are exactly periodic $\{0,1\}$-valued sequences over $\mathbb{Z}$. Let

$$\bar{d}(A) = \limsup_{n\to\infty} \frac{|A \cap \{0,1,\dots,n-1\}|}{n}.$$

For $A, B \subset \mathbb{Z}$ the formula $\bar{d}(A \div B)$ introduces a pseudometric on $\mathscr{P}(\mathbb{Z})$. Following Bergelson and Ruzsa, we call a set $A \subset \mathbb{Z}$ *rational* if it belongs to the $\bar{d}$ closure of the family of periodic sets.

Our motivating examples are sets of $\mathscr{B}$-free integers. An integer is *$\mathscr{B}$-free* if it has no factor in a given set $\mathscr{B} \subset \mathbb{N}$. For example, the set of $\mathscr{B}_{\mathrm{sq}}$-free integers where $\mathscr{B}_{\mathrm{sq}} = \{p^2 : p \text{ prime}\}$ is just the set of *square-free integers*. These sets were studied by Chowla, Davenport and Erdős. Recently, Sarnak [7] initiated the study of a symbolic dynamical system X_{sq} associated with $\mathscr{B}_{\mathrm{sq}}$. Abdalauoi, Lemańczyk, and de la Rue [1] extended Sarnak's approach to $\mathscr{B}$-free systems determined by any $\mathscr{B}$ consisting of infinitely many pairwise relatively prime integers, the sum of whose reciprocals is finite (we call such a set $\mathscr{B}$ an *Erdős set*). Bartnicka et al. [2] considered shift spaces associated with $\mathscr{B}$-free integers for an arbitrary $\mathscr{B}$. These systems were also investigated by Cellarosi and Sinai [3], Kułaga-Przymus, Lemańczyk, and Weiss [4, 5], Peckner [6].

Writing elements of $\mathscr{B} \subset \mathbb{N}$ as an increasing sequence $b_1, b_2, \dots$ we may consider the periodic set $\mathscr{F}_{\mathscr{B}}^{(k)}$ of $\mathscr{B}^{(k)}$-free integers (here $\mathscr{B}^{(k)} = \{b_1, \dots, b_k\}$) as a periodic approximation on $\mathscr{F}_{\mathscr{B}}$. Indeed, it is often the case that $\mathscr{F}_{\mathscr{B}}$ is rational, that is $\bar{d}(\mathscr{F}_{\mathscr{B}} \div \mathscr{F}_{\mathscr{B}}^{(k)}) \to 0$ as $k \to \infty$. Using the Davenport-Erős theorem we formalize the vague statement that $\mathscr{F}_{\mathscr{B}} \subset \mathbb{Z}$ is approximated by $\mathscr{F}_{\mathscr{B}}^{(k)}$ for any $\mathscr{B}$ and hence we generalize the notion of a rational set.

A natural symbolic dynamical system (aka a *shift space*) associated with a set $A \subset \mathbb{Z}$ arises from the identification of A with its characteristic sequence $\boldsymbol{a}$. Then $\boldsymbol{a}$ is a biinfinite $\{0,1\}$-valued sequence, which is a point in the space $\Omega = \{0,1\}^{\mathbb{Z}}$, where $\{0,1\}$ is given the discrete topology and Ω is given the corresponding product topology. The left-shift operator $\sigma\colon \Omega \to \Omega$ is a homeomorphism of Ω and the closure $X_{\boldsymbol{a}}$ of the σ-orbit of is a shift space (it is closed, nonempty and σ-invariant). This construction was used by Furstenberg in his proof of the Szemerédi theorem. We say that $X_{\boldsymbol{a}}$ is a *rational shift* if A is a rational set. The *$\mathscr{B}$-free shift* $X_{\boldsymbol{b}}$ is the closure of the orbit of $\boldsymbol{b} \in \Omega$, where $\boldsymbol{b}$ is the characteristic function of the set of $\mathscr{B}$-free integers $\mathscr{F}_{\mathscr{B}} \subset \mathbb{Z}$.

Another shift space connected with $\mathscr{F}_{\mathscr{B}}$ is the *$\mathscr{B}$-admissible shift* $X_{\mathscr{B}}$. It consists of all $\mathscr{B}$-admissible sequences in Ω, where we say that $x = (x_j)_{j\in\mathbb{Z}} \in \Omega$ is *$\mathscr{B}$-admissible* if for every $b \in \mathscr{B}$ the set $\{j \in \mathbb{Z} : x_j = 1\}$ is disjoint with a set $b\mathbb{Z} + r$ for some $0 \le r < b$. Since $\boldsymbol{b}$ is clearly a $\mathscr{B}$-admissible sequence we see immediately that $X_{\boldsymbol{b}} \subset X_{\mathscr{B}}$. Furthermore, if $\mathscr{B}$ is an Erdős set, then $X_{\boldsymbol{b}} = X_{\mathscr{B}}$ by [1, Cor. 2.6]. This is not the case in general and is a reason to introduce yet another construction. A shift space over $\{0,1\}$ is *hereditary* if it is closed with respect to a coordinatewise

Date: September 30, 2016.

multiplication by an arbitrary 0-1 sequence. Given a shift space X by $\tilde{X}$ we denote the *hereditary closure of* X which is the smallest hereditary shift space containing X. Note that the $\mathscr{B}$-admissible shift is hereditary for every $\mathscr{B}$, hence $\tilde{X}_{\mathscr{B}} = X_{\mathscr{B}}$ and we always have $X_{\boldsymbol{b}} \subset \tilde{X}_{\boldsymbol{b}} \subset X_{\mathscr{B}}$.

We study invariant measures for hereditary closures of $\mathscr{B}$-free and rational shifts. We show that ergodic invariant measures of any shift space in that family are abundant and their structure resemble invariant measures of a transitive uniformly hyperbolic system: *The ergodic measures are entropy dense*. That is, any invariant measure μ is a weak*limit of a sequence of ergodic measures with Kolmogorov-Sinai entropies also converging to the entropy of μ. This extends results from [5, 2], where density, but not ergodic density of ergodic measures is proved for all hereditary closures of $\mathscr{B}$-free shifts in two directions: we add entropy to the picture and broaden the class of shift spaces for which this result holds. For our purposes we develop techniques of *single orbit dynamics* and heavily use $\bar{d}$-pseudometric and related concepts of independent interest. As a matter of fact these methods lead to new, often shorter proofs, of many results from [4, 5], and [2] in our more general setting. We also describe the ways in which a subset of $\mathbb{Z}$ is approximated by periodic sets which leads to many intriguing questions of the combinatorial number theoretic properties of theses sets.

References

[1] H. Abdalauoi, M. Lemańczyk, T. de la Rue, *A dynamical point of view on the set of $\mathscr{B}$-free integers*. Preprint, `arXiv:1311.3752`, 2013.

[2] A. Bartnicka, S. Kasjan, J. Kułaga-Przymus, and M. Lemańczyk, *$\mathscr{B}$-free sets and dynamics*, Preprint, 2015.

[3] F. Cellarosi, Y. G. Sinai, *Ergodic Properties of Square-Free Numbers*. Journal of the European Mathematical Society, **15** (2013) , 1343-1374.

[4] J. Kułaga-Przymus, M. Lemańczyk, B. Weiss, *On invariant measures for $\mathscr{B}$-free systems*. Proc. Lond. Math. Soc. (3), 110, 1435-1474, 2015.

[5] J. Kułaga-Przymus, M. Lemańczyk, B. Weiss, *Hereditary subshifts whose simplex of invariant measures is Poulsen*. Preprint `arXiv:1507.00714[math.DS]`, 2015.

[6] R. Peckner *Uniqueness of the measure of maximal entropy for the squarefree flow*. Preprint `arXiv:1205.2905[math.DS]`, to appear in Israel J. Math, 2014.

[7] P. Sarnak. *Three lectures on the Möbius function randomness and dynamics (Lecture 1)*. `http://publications.ias.edu/sites/default/files/MobiusFunctionsLectures(2).pdf`.

[8] P. Shields. *The Ergodic Theory of Discrete Sample Path*. Vol. 13. American Mathematical Society, 1991. Shields,

[9] Benjamin Weiss, *Single orbit dynamics*, CBMS Regional Conference Series in Mathematics, vol. 95, American Mathematical Society, Providence, RI, 2000. MR 1727510 (2000k:37001)

(D. Kwietniak) Faculty of Mathematics and Computer Science, Jagiellonian University in Krakow, ul. Łojasiewicza 6, 30-348 Kraków, Poland and Institute of Mathematics, Federal University of Rio de Janeiro, Cidade Universitaria - Ilha do Fundão, Rio de Janeiro 21945-909, Brazil

E-mail address: `dominik.kwietniak@uj.edu.pl`

URL: `www.im.uj.edu.pl/DominikKwietniak/`

(M. Kupsa) Institute of Information Theory and Automation, The Academy of Sciences of the Czech Republic, Prague 8, CZ-18208

E-mail address: `m@cz`

(J. Konieczny) Mathematical Institute, University of Oxford, Andrew Wiles Building, Radcliffe Observatory Quarter, Woodstock Road, Oxford, OX2 6GG

E-mail address: `jakub.konieczny@gmail.com`

On arithmetic index in the Thue-Morse word

Parshina O. G.

Sobolev Institute of mathematic SB RAS, Russia

Université de Lyon, Université Lyon 1, Institut Camille Jordan, France

Let w be an infinite word over the finite alphabet Σ, $|\Sigma| = t$. A set $A_w(d) = \{w_c w_{c+d} w_{c+2d} \cdots w_{c+(k-1)d} | c, k \in \mathbb{Z}^+\}$ defined for every positive integer d is the set of arithmetic subsequences in w obtained with the difference d. If the arithmetic closure of w: $\cup_{d\in\mathbb{Z}^+} A_w(d)$ contains all finite words, the word w is called arithmetic universal (AU-word)[1]. A well-known example of a AU-word is the Thue-Morse word [2] defined as $w_{TM} = w_0 w_1 w_2 w_3 \cdots$, where $w_i \in \{0, 1\}$ is the sum modulo 2 of digits in the binary representation of i.

Consider a word u of length n over Σ and define a number $i_w(u) = \min_d\{u \in A_w(d)\}$. The word u has an arithmetic index $I_w(u)$ if $I_w(u)$ is the length of the t-ary notation of $i_w(u)$. In the case u does not belong to the arithmetic closure of w we set $I_w(u) = \infty$. The object of the research is a function $\max_{u:|u|=n} I_w(u)$.

The function $\max_{u:|u|=n} I_{w_{TM}}(u)$ for the Thue-Morse word is considered. By this moment, a lower and an upper bounds on the rate of growth of this function have been obtained. A lower bound is based on the formula of the factor complexity of w_{TM} [3] and is equal to $(n - \log n - 2)/2$. An upper bound is based on the theorem about a distribution of arithmetic progressions – arithmetic subsequences consisting of the same symbols – in the Thue-Morse word formulated below.

Consider a function $a(c, d)$ which outputs the length of an arithmetic progression with starting symbol v_c and difference d for positive integers c and d in the Thue-Morse word. The function $a(d) = \max_c a(c, d)$ outputs the length of the maximal arithmetic progression with the difference d.

Theorem 1 *For all numbers $n \geq 1$ the following holds:*

$$\max_{d<2^n} a(d) = \begin{cases} 2^n + 4, & n \equiv 0 \ mod\ 2, \\ 2^n, & otherwise. \end{cases}$$

A similar result for the generalized Thue-Morse word over the ternary alphabet one can find in [4].

A natural corollary from the theorem 1 is that the arithmetic index of every binary word of the form 0^n or 1^n is not greater than $\log n$. The upper bound for arbitrary binary word of length n is $3n \log n$.

Computer experiments have been carried out for binary words of length $n \leq 18$. According to experiments the lower bound is closer to the real growth of considered function.

References

[1] Avgustinovich S. V., Fon-der-Flaass D. G., Frid A. E. Arithmetical complexity of infinite words // Proc. Words, Languages and Combinatorics III, 2000. Singapore: World Scientific, 2003, P. 51-62.

[2] Thue A. Uber die gegenseitige Lage gleicher Teile Gewisser Zeichenreichen //Skr. Vid.-Kristiana I. Mat. Naturv. Klasse — 1912 — Vol.1 — P. 1–67

[3] Avgustinovich S. V. The number of different subwords of fixed length in the Morse-Hedlund sequence [Russian] // Siberian Journal of Operational Research — T.1, Vol.2 — 1994 — P. 3-7

[4] Parshina O. G. On arithmetic progressions in the generalized Thue-Morse word // Lecture Notes in Computer Science 9304 — 2015 — P. 191-196

Palindromic Length in Linear Time

Mikhail Rubinchik, Arseny M. Shur
Ural Federal University
Ekateinburg, Russia

Palindromes are one of the most important repetitive structures in strings. During the last decades they were actively studied in formal language theory, combinatorics on words and stringology. Recall that a palindrome is any string $S = a_1a_2 \cdots a_n$ equal to its reversal $a_n \cdots a_2a_1$. There is a lot of papers concerning the palindromic structure of strings. The most important problems in this direction include the search and counting of palindromes in a string and the factorization of a string into palindromes.

There are two versions of the palindromic factorization problem. In the k-factorization problem, it is required to factorize a string into a fixed number k of palindromes or establish that no such factorization exists (this can be viewed as recognizing the language Pal^k). The palindromic length problem asks to factorize a string into the minimal number of palindromes. As was shown in [1], k-factorization can be solved in time $O(kn)$. Another algorithm for k-factorization, presented in [3], works in $O(n \log n)$ time independently of k. There are two solutions [2, 4] of palindromic length problem in $O(n \log n)$ time. In [3] we presented more practical algorithm with the same complexity. All the mentioned algorithms are online, that is, they give the answer for each prefix of the input string before reading the next symbol. The main question is, are there any faster algorithms? This was open until now. In this talk we present a linear time online algorithm for the palindromic length problem.

The algorithm uses the bit compression technique (so-called "four Russians' trick") applied to a specific representation of the solution: instead of an array of n integers to store the palindromic lengths of all prefixes of the processed string, we use only $2n$ bits to store the differences between the consecutive elements of this array. The main procedure is the computation of the new list of suffix-palindromes from the old list after appending a new symbol; this procedure inevitably takes $O(\log n)$ time. We show how to apply this procedure only $O(n/\log n)$ times, reducing all other time expenses per iteration to $O(1)$. This allows us to get the overall linear time bound.

References

[1]

[2] G. Fici, T. Gagie, J. Kärkkäinen, D. Kempa. A subquadratic algorithm for minimum palindromic factorization // J. of Discrete Algorithms. — 2014. — Vol. 28. — P. 41–48.

Kosolobov D., Rubinchik M., Shur A. M. Pal^k is linear recognizable online // SOFSEM 2015: Theory and Practice of Computer Science. — Springer-Verlag Berlin Heidelberg, 2015. — Vol. 8939 of LNCS. — P. 289–301.

[3] Rubinchik Mikhail, Shur Arseny M. EERTREE: An Efficient data structure for processing palindromes in strings // Combinatorial algorithms: Proc. IWOCA 2015. — Vol. 9538 of LNCS. — Springer International Publishing, 2016. — P. 321–333.

[4] Tomohiro I, Shiho Sugimoto, Shunsuke Inenaga, Hideo Bannai, Masayuki Takeda. Computing palindromic factorizations and palindromic covers on-line //Symposium on Combinatorial Pattern Matching. – Springer International Publishing, 2014. – Pages. 150-161.

Synchronization of Weakly Acyclic Automata

Andrew Ryzhikov

Grenoble Alpes University, France

ryzhikov.andrew@gmail.com

The concept of synchronizing automata is widely studied in automata theory and has a lot of different applications in such areas as manufacturing, biocomputing, semigroup theory and many others [Vol08]. Let $A = (Q, \Sigma, \delta)$ be a deterministic finite automaton, where Q is the set of its states, Σ is a finite alphabet and $\delta : Q \times \Sigma \to Q$ is a transition function. Note that our definition of automata does not include initial and accepting states. An automaton is called *synchronizing* if there exists a word that maps every its state to a fixed state $q \in Q$. A set S of states of an automaton A is called *synchronizing* if there exist some word $w \in \Sigma^*$ and some state $q \in Q$ such that after reading the word w starting in any state $s \in S$, A ends up in q. The word w is said to *synchronize* the set S.

An automaton is called *binary* if its alphabet has size two. A *cycle* in an automaton is a sequence $q_1, \ldots, q_n$ of its states such that there exist letters $x_1, \ldots, x_n \in \Sigma$ with $\delta(q_i, x_i) = q_{i+1}$ for $1 \leq i \leq n-1$ and $\delta(q_n, x_n) = q_1$. A cycle is a *self-loop* if it consists of one state. An automaton is called *weakly acyclic* if all its cycles are self-loops. Weakly acyclic automata are called acyclic in [JM12] and partially ordered in [BF80]. Weakly acyclic automata arise naturally in the synchronizing automata theory. For example, the automata in the reductions proving the facts that the problem of finding a shortest synchronizing word is NP-complete for binary automata [Epp90] and hard to approximate for automata with alphabet of non-constant size [Ber14] are weakly acyclic.

One of the most important questions in the synchronizing automata theory is the famous Černý conjecture stating that any n-state synchronizing automaton has a synchronizing word of length at most $(n-1)^2$. For weakly acyclic automata, we prove a stronger property. Given an automaton A, the *rank* of a word w is the number $|\{\delta(s, w) \mid s \in Q\}|$.

Theorem 1. Let A be a synchronizing weakly acyclic automaton, and w be a word of rank r with respect to A. Then there exists a word of length $n-r$ and rank r with respect to A.

The problem of deciding whether the given automaton is synchronizing is solvable in polynomial time [Vol08]. However, the problem of deciding whether the set S of states of a given automaton A is synchronizing is PSPACE-complete (see [San05] for a survey on this subject), even for binary strongly connected automata. For weakly acyclic automata, we prove that the following results hold.

Theorem 2. Let S be a synchronizing set of states in a weakly acyclic n-state automaton A. Then the length of a shortest word synchronizing S is at most $\frac{n(n-1)}{2}$.

Corollary 1. The Sync State problem for weakly acyclic automata is in NP.

Theorem 3. The Sync Set problem is NP-complete for binary weakly acyclic automata.

Next, we introduce a related problem Max Sync Set of finding a synchronizing set of states of maximum size in a given automaton. We investigate the complexity and approximability of

this problem. The proofs of the statements can be found in the pre-print on arXiv [Ryz16].

Theorem 4. The decision version of the Max Sync Set problem is PSPACE-complete for binary automata.

Theorem 5. The problem Max Sync Set for weakly acyclic n-state automata over an alphabet of cardinality $O(n)$ cannot be approximated in polynomial time within a factor of $O(n^{1-\varepsilon})$ for any $\varepsilon > 0$ unless P = NP.

Theorem 6. The Max Sync Set problem for binary n-state automata cannot be approximated in polynomial time within a factor of $O(n^{\frac{1}{2}-\varepsilon})$ for any $\varepsilon > 0$ unless P = NP.

Theorem 7. The Max Sync Set problem for binary weakly acyclic n-state automata cannot be approximated in polynomial time within a factor of $O(n^{\frac{1}{3}-\varepsilon})$ for any $\varepsilon > 0$ unless P = NP.

Finally, we use the developed technique to show the inapproximability of the problem of computing the rank of a subset of states. A *rank* of a set of states $S \subseteq Q$ is the number $\min_{w\in\Sigma^*} |\{\delta(s,w) \mid s \in S\}|$. This notion generalizes the rank of an automaton, which is the rank of the set of all states of an automaton. The rank of an automaton can be computed in polynomial time [Rys92].

Theorem 8. The problem Set Rank of computing the rank of a given subset of states in an n-state binary weakly acyclic automaton cannot be approximated in polynomial time within a factor of $O(n^{\frac{1}{4}-\epsilon})$ for any $\epsilon > 0$ unless P = NP.

References

[Ber14] Mikhail V. Berlinkov. On two algorithmic problems about synchronizing automata. In Arseny M. Shur and Mikhail V. Volkov, editors, *Developments in Language Theory: 18th International Conference, DLT 2014, Ekaterinburg, Russia, August 26-29, 2014. Proceedings*, pages 61–67. Springer International Publishing, Cham, 2014.

[BF80] J.A. Brzozowski and Faith E. Fich. Languages of R-trivial monoids. *Journal of Computer and System Sciences*, 20(1):32 – 49, 1980.

[Epp90] David Eppstein. Reset sequences for monotonic automata. *SIAM Journal on Computing*, 19(3):500–510, 1990.

[JM12] Galina Jirásková and Tomáš Masopust. On the state and computational complexity of the reverse of acyclic minimal DFAs. In Nelma Moreira and Rogério Reis, editors, *Implementation and Application of Automata: 17th International Conference, CIAA 2012, Porto, Portugal, July 17-20, 2012. Proceedings*, pages 229–239. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.

[Rys92] Igor K. Rystsov. Rank of a finite automaton. *Cybernetics and Systems Analysis*, 28(3):323–328, 1992.

[Ryz16] Andrew Ryzhikov. Approximating the Maximum Number of Synchronizing States in Automata. *ArXiv e-prints, 1608.00889*, August 2016.

[San05] Sven Sandberg. Homing and synchronizing sequences. In Manfred Broy, Bengt Jonsson, Joost-Pieter Katoen, Martin Leucker, and Alexander Pretschner, editors, *Model-Based Testing of Reactive Systems: Advanced Lectures*, pages 5–33. Springer Berlin Heidelberg, Berlin, Heidelberg, 2005.

[Vol08] Mikhail V. Volkov. Synchronizing automata and the Černý conjecture. In Carlos Martín-Vide, Friedrich Otto, and Henning Fernau, editors, *Language and Automata Theory and Applications: Second International Conference, LATA 2008, Tarragona, Spain, March 13-19, 2008. Revised Papers*, pages 11–27. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

Maximal Edit Distance to a Synchronizing Coloring of a Graph

This research is focused on k-out-regular directed multigraphs on n vertices with loops (called simply *digraphs*). The edges of such a digraph can be colored by elements of some fixed k-element set in such a way that outgoing edges of every vertex have different colors. Such a coloring corresponds naturally to an automaton.

In 1977 Adler, Goodwyn and Weiss conjectured [1] that every primitive digraph has a synchronizing coloring. This conjecture became widely known as the road coloring problem and was proved by Trahtman in 2007 [3].

The main motivation for this work comes from the algorithmic issues related to the road coloring problem. How to find a synchronizing coloring of a given digraph? Trahtman's proof provides an algorithm for this task working in time $O(n^3)$. Later his construction was improved [2], providing a non-trivial algorithm with a worst-case complexity $O(kn^2)$.

Both of this algorithms construct a synchronizing coloring by taking a random coloring and successively changing it, until it becomes synchronizing. The natural questions arise — what is the maximal number of such changes is needed, to make any given coloring synchronizing and on what kinds of digraphs it can be achieved?

Let $\varrho(\mathcal{G})$ be the maximal number of edits required to make a synchronizing coloring from any coloring of digraph $\mathcal{G}$.

In my research, an experimental and theoretical study on the maximal values of such edit distance for graphs on n vertices is performed. The main results are as follows:

1. Developed an efficient algorithm for enumerating non-isomorphic digraphs and computing $\varrho(\mathcal{G})$.

2. Using this algorithm for small n and k, an extensive experiments are performed, providing evidence to state a conjecture, giving the upper bound of $\varrho(\mathcal{G}) \leq \log_2 n$.

3. The series of graphs and their colorings achieving

   a. $\varrho(\mathcal{G}^1_m) = m = \log_2 n$, for $n = 2^m$ and $k = m + 1$.
   The digraph $\mathcal{G}^1_m$ is Cayley graph of a group with generating set $\{id; (1,2); (3,4); \ldots; (2m-1, 2m)\}$.

b. $\varrho(\mathcal{G}_m^2) = m+1$, for $n = 2 \cdot 3^m$ and $k = m+1$.
The digraph $\mathcal{G}_m^2$ is Cayley graph of a group with generating set $\{(1,2,3);(4,5,6);\ldots;(3m-2,3m-1,3m);$
$(1,2)(4,5)\ldots(3m-2,3m-1)\}$.

c. $\varrho(\mathcal{G}_m^3) = m-1$, for $n = m!$ and $k = 2$.
The digraph $\mathcal{G}_m^3$ is Cayley graph of the symmetric group S_m with generating set $\{(1,2,\ldots,n-1),(1,2,\ldots n)\}$

References

[1] R. L. Adler, L. W. Goodwyn, and B. Weiss. Equivalence of topological markov shifts. *Israel Journal of Mathematics*, 27(1):49–63, 1977.

[2] M.-P. Béal and D. Perrin. A quadratic algorithm for road coloring. *Discrete Applied Mathematics*, 169:15–29, 2014.

[3] A. N. Trahtman. The road coloring problem. *Israel Journal of Mathematics*, 172(1):51–60, 2009.