Nonconstructive methods for nonrepetitive problems

Guillaume Guégan*and Pascal Ochem†

May 2014

Abstract

We obtain the following three results: (1) There is a strategy to build an arbitrarily long nonrepetitive word over 6 letters in the erase-repetition game. (2) From every assignment of lists of size 5 to the vertices of a caterpillar graph, we can extract a nonrepetitive coloring. (3) There exist infinite words avoiding shuffle squares over 7 letters. Results (1) and (2) are obtained using the entropy compression method. Result (3) is obtained by the power series technique of Bell and Goh.

1 Introduction

Grytczuk *et al.* [5] introduced the *erase-repetition game* over a finite alphabet. It is a two-player game between Ann and Ben. They build a word by alternately choosing a letter and appending it to the end of the current word. Whenever a square occurs, the second half of the square is erased and the next player continues extending the remaining prefix of the sequence. Ann's goal is to obtain an arbitrarily long (squarefree) word whereas Ben's goal is to prevent this. Grytczuk *et al.* [5] have obtained a winning strategy for Ann with an alphabet of size 8. We lower this bound to 6 in Section 2.

It is known that the nonrepetitive chromatic number of the class of caterpillars is 4. In Section 3, we show that the nonrepetitive choice number of the class of caterpillars is 4 or 5.

Recently, Currie [2] has answered a question of Karhumäki by showing that shuffle squares (see [2] for the definition) are avoidable over an alphabet of size $\lceil e^{115} \rceil$ using the Lovász local lemma. We lower this bound to 7 in Section 4.

For the first two results, we use the entropy compression method as described in [5] to prove that the nonrepetitive choice number of the path is at most 4. The last result is also nonconstructive and uses power series.

2 The erase-repetition game

Theorem 1. *In the erase-repetition game over an alphabet of size 6, there exists a strategy for Ann to build an arbitrarily long squarefree word.*

Proof. We describe Ann's strategy. Let $w = w_1 \dots w_i$ be the word before Ann's turn. If $w_{i-3} = w_i$ then Ann chooses a letter that is distinct from w_i, w_{i-1}, and w_{i-2}. Otherwise, let j be the largest integer such that $j < i$, $w_{j-1}w_j = w_{i-1}w_i$, and $w_{j+1} \neq w_{i-1}$. Then Ann

*Corresponding author, Univ. Montpellier 2, LIRMM

†CNRS, LIRMM

chooses a letter distinct from w_i, w_{i-1}, and w_{j+1} (if j is defined). This strategy implies that Ann and Ben do not generate repetitions of size at most 11. The proof of this assertion is omitted due to lack of space.

Suppose there exists n such that Ben has a strategy to keep the length of the built word smaller than n. Let t be the number of moves for Ann. We say that a move of Ben is *trivial* if he adds to the current word its last letter, so that the added letter is immediatly erased. We record the final built word and the partial Dyck word corresponding to the history of the length of the built word, except that we omit Ben's trivial moves. In this Dyck word, all the descents have length at least 12. There are $o(1.485^k)$ Dyck words of length $2k$ such that all the descents have length at least 12 (see [3] for the computation of the asymptotics of Dyck words with a constrained set of descent lengths). For $k = 2t$, this gives $7^n \times o(1.485^{2t}) = o(3^t)$ distinct records, whereas there are 3^t possible executions. This is a contradiction since we can recover Ann's moves from Ben's moves, the final built word, and the partial Dyck word, as proved in [5].

An alphabet of size 6 is thus sufficient for Ann's strategy: at each step, Ann discards three letters to ensure that Ann and Ben cannot create a non-trivial square of size at most 11, then the three letters that remain are sufficient for the entropy compression argument. □

3 Nonrepetitive list coloring of caterpillars

It is known that trees admit a nonrepetitive coloring with 4 colors [6] and it is easy to check that 4 colors are indeed needed for a sufficiently long caterpillars of maximum degree 3. For the list version of nonrepetitive coloring, the corresponding choice number is unbounded for trees [4] but is bounded for graphs of bounded pathwidth [7]. In this section we consider nonrepetitive list coloring of caterpillars, i.e., graphs with pathwidth 1, and we prove the following:

Theorem 2. *The nonrepetitive choice number of a caterpillar is at most 5.*

Proof. We first color the vertices of the spine using entropy compression. We require that the word corresponding to the colors on the spine avoids squares and factors of the form $AxyA$ where x and y are letters and A is a non-empty word. We suppose that there exists a path of length n and lists of size 5 on its vertices such that we cannot extract a coloring avoiding squares and factors $AxyA$. The algorithm proceeds as follows: If the suffix of the current word is aba then we choose a color distinct from a and b, otherwise this suffix is abc with $a \neq c$ and we choose a color distinct from a and c. We thus discard two colors in order to forbid factors of the form aa, $abab$, and $abca$. If the color chosen by the algorithm creates a factor $AxyA$, then we erase the suffix of length $|A|$. Since the factors $abca$ are forbidden, we have that $|A| \geqslant 2$. After t steps, the algorithm has recorded a partial coloring and a partial Dyck word corresponding to the number of colored vertices during the execution of the algorithm. Since we never erase exactly one letter, the Dyck word has no descent of length 1. It is well-known that the number of such partial Dyck words is $o(3^t)$. The number of potential records is at most $5^n \times o(3^t) = o(3^t)$, whereas there are 3^t possible executions (see [3] for the computation of the asymptotics of Dyck words with a constrained set of descent lengths).

Lists of size 5 are thus sufficient to avoid squares and factors $AxyA$: two colors are discarded in order to forbid aa, $abab$, and $abca$, then the remaining list of size at least 3 is able to avoid the factors $AxyA$ with $|A| \geqslant 2$.

Now we color the leaves (vertices of degree 1) with a color distinct from the colors of the 3 vertices at distance at most two that belong to the spine. Since the size of the lists is 5, there remain at least 2 possibilities to color a leaf.

To finish the proof, we check that coloring the leaves does not create a square. Squares of size at most two are avoided by the previous rule. A square of size at least three would imply a factor $AxyA$ in the coloring of the spine. □

4 Avoiding shuffle squares

The following theorem was originally presented by Golod (see [9], Lemma 6.2.7) and rewritten and proven with combinatorial terminology by Rampersad [8].

Theorem 3. *Let S be a set of words over an m-letter alphabet, each word of length at least 2. Suppose that for each $i \geqslant 2$, the set S contains at most a_i words of length i. If the power series expansion of*

$$G(x) := \left(1 - mx + \sum_{i \geqslant 2} a_i x^i\right)^{-1} \tag{1}$$

has non-negative coefficients, then there are least $[x^n]G(x)$ words of length n over an m-letter alphabet that avoid S.

To use Theorem 3, we need to obtain reasonable upper bounds on the number a_i of forbidden factors. These forbidden factors are the *minimal shuffle squares*, i.e., the shuffle squares that do not contain a smaller shuffle square. To a shuffle square S of a word w of length i, we associate the *height function* $h\colon [0, \dots, 2i] \to \mathbb{Z}$ defined as follows:

- $h(0) = 0$.
- For $0 < j \leqslant 2i$ if $S[j]$ belongs to the left factor w, then $h(j) = h(j-1) + 1$.
- For $0 < j \leqslant 2i$ if $S[j]$ belongs to the right factor w, then $h(j) = h(j-1) - 1$.

Notice that $h(2i) = 0$. Moreover, if $h(j) = 0$ for some $0 < j < 2i$, then the prefix of length j of S is a shuffle square. So, if h is the height function of a minimal shuffle square, then $h(2i) > 0$ for every $0 < j < 2i$. Thus, there is a bijection between height functions of minimal shuffle squares and Dyck words of length $2i - 2$. The number of such functions is thus $C_{i-1} = \frac{(2i-1)!}{i!\,(i-1)!}$. We consider the m-letter alphabet. A minimal shuffle square S of a word w of length i is determined by w and the height function of S. There are thus at most $C_{i-1}m^i$ minimal shuffle squares of length $2i$. We will need sharper bounds on the number of small minimal shuffle squares: there are $m(m-1)$ minimal shuffle squares of length 4 and $2m(m-1)(m-2)$ minimal shuffle squares of length 6.

To every word w of length $i \geqslant 1$ over $\Sigma_m = \{0, \dots, m-1\}$ avoiding (shuffle) squares of length 2, we associate a *code* c (similar to Pansiot's code) of length $i - 1$ over the $m - 1$ letters $\{1, \dots, m-1\}$ such that $c[j] = (w[j+1] - w[j]) \pmod{m}$ for $1 \leqslant j < i$. Notice that exactly m words w (determined by their first letter) have the same code. By the previous observations, we can upper bound the number of codes of minimal shuffle squares:

- There are $(m-1)$ codes of length 3 corresponding to minimal shuffle squares of length 4.
- There are $2(m-1)(m-2)$ codes of length 5 corresponding to minimal shuffle squares of length 6.
- There are at most $C_{i-1}m^{i-1}$ codes of length $2i-1$ corresponding to minimal shuffle squares of length $2i$.

Now, we can apply Theorem 3 to the language of codes of words avoiding shuffle squares. Thus we consider the power series expansion of

$$G(x) := \left(1-(m-1)x+(m-1)x^3+2(m-1)(m-2)x^5+\sum_{i\geqslant 4} C_{i-1}m^{i-1}x^{2i-1}\right)^{-1}$$

We set $m=7$ and $s_k=[x^k]G(x)$. We prove by induction on k that $s_k > \sqrt{30}s_{k-1}$. From the relation

$$1=\left(1-6x+6x^3+60x^5+\sum_{i\geqslant 4} C_{i-1}7^{i-1}x^{2i-1}\right)\left(1+\sum_{k\geqslant 1} s_k x^k\right),$$

we deduce

$$s_k = 6s_{k-1}-6s_{k-3}-60s_{k-5}-\sum_{4\leqslant j\leqslant (k+1)/2} C_{j-1}7^{j-1}s_{k-2j+1} \text{ for } k\geqslant 1.$$

By induction, we have $s_{k-2j+1} < \frac{s_{k-1}}{(\sqrt{30})^{2j-2}} = \frac{s_{k-1}}{30^{j-1}}$ for $1\leqslant j\leqslant (k+1)/2$.

We obtain

$$\begin{aligned}
s_k &= 6s_{k-1}-6s_{k-3}-60s_{k-5}-\sum_{4\leqslant j\leqslant (k+1)/2} C_{j-1}7^{j-1}s_{k-2j+1}\\
&> 6s_{k-1}-\frac{6}{30}s_{k-1}-\frac{60}{30^2}s_{k-1}-\sum_{4\leqslant j\leqslant (k+1)/2} C_{j-1}7^{j-1}\frac{s_{k-1}}{30^{j-1}}\\
&= \left(\frac{86}{15}-\sum_{4\leqslant j\leqslant (k+1)/2} C_{j-1}\left(\frac{7}{30}\right)^{j-1}\right)s_{k-1}\\
&= \left(\frac{86}{15}-\sum_{3\leqslant j\leqslant (k-1)/2} C_{j}\left(\frac{7}{30}\right)^{j}\right)s_{k-1}\\
&> \left(\frac{86}{15}-\sum_{j\geqslant 3} C_{j}\left(\frac{7}{30}\right)^{j}\right)s_{k-1}\\
&= \left(\frac{86}{15}+1+\frac{7}{30}+2\left(\frac{7}{30}\right)^2-\sum_{j\geqslant 0} C_{j}\left(\frac{7}{30}\right)^{j}\right)s_{k-1}
\end{aligned}$$

Notice that $\sum_{j\geqslant 0} C_j z^j = \frac{2}{1+\sqrt{1-4z}}$ for $|z|<\frac{1}{4}$, since it corresponds to the power series of Catalan numbers. We thus obtain

$$s_k > \left(\frac{1592}{225}-\frac{2}{1+\sqrt{1-4\times\frac{7}{30}}}\right)s_{k-1} > \sqrt{30}s_{k-1},$$

which proves the induction. This shows that $s_k \geqslant 30^{k/2}$. By Theorem 3, there exist at least $30^{n/2}$ codes of length n. This means that there exist at least $7\times 30^{(n-1)/2}$ words of length $n\geqslant 1$ avoiding shuffle squares over a 7-letter alphabet. We have thus proved the following:

Theorem 4. *Shuffle squares are 7-avoidable.*

References

[1] J. Bell, T. L. Goh. Exponential lower bounds for the number of words of uniform length avoiding a pattern, *Inform. and Comput.* **205** (2007), 1295-1306.

[2] J. Currie. Shuffle squares are avoidable.

[3] L. Esperet and A. Parreau. Acyclic edge-coloring using entropy compression. *European Journal of Combinatorics* **34(6)** (2013), 1019–1027.

[4] F. Fiorenzi, P. Ochem, P. Ossona de Mendez, and X. Zhu. Thue choosability of trees. *Discrete Applied Mathematics* **159** (2011), 2045–2049.

[5] J. Grytczuk, J. Kozik, and P. Micek. A new approach to nonrepetitive sequences, *Random Structures & Algorithms* **42(2)** (2013), 214–225.

[6] A. Kündgen and M. J. Pelsmajer. Nonrepetitive colorings of graphs of bounded tree-width. *Discrete Mathematics* **308(19)** (2008), 4473–4478.

[7] P. Micek. Personnal communication.

[8] N. Rampersad. Further applications of a power series method for pattern avoidance. *Electron. J. Combinatorics.* **18(1)** (2011), #P134.

[9] L. Rowen, *Ring Theory. Vol. II*, Pure and Applied Mathematics 128, Academic Press, Boston, 1988.