Upper bound on the number of ternary square-free words

Pascal Ochem and Tony Reix

LaBRI, Université Bordeaux 1
351 cours de la Libération
33405 Talence Cedex, France
ochem@labri.fr
Tony.Reix@bull.net

Abstract

Let t_n be the number of words of length n in a factorial language L. We adapt the transfert matrix method to obtain upper bounds on the growth rate of words in L defined as $\lim_{n\to\infty}(t_n)^{\frac{1}{n}}$. This method is used to lower the best known upper bounds on the growth rate of ternary square-free words from 1.30193812.. to 1.30178858..

1 Introduction

Let Σ_s denote the s-letter alphabet $\{0, 1, \ldots, s-1\}$. A language L is said to be *factorial* if it is closed under taking factors (i.e., subwords consisting of consecutive letters). Let t_n be the number of words of length n in a factorial language L over Σ_s. Since L is factorial, t_n is sub-multiplicative and thus the growth rate $C = \lim_{n\to\infty}(t_n)^{\frac{1}{n}}$ is well-defined. We are in one the following case:

- If L contains finite words only, then $C = 0$.
- If $L = \Sigma_s^*$, then $C = \lim_{n\to\infty}(s^n)^{\frac{1}{n}} = s$.
- Otherwise $1 \leq C < s$.

See the survey of Berstel [1] for more information on the growth rate.

A factorial langage can also be defined by its set of minimal forbidden factors. If this set is finite, then we can obtain the generating function of t_n and compute the growth rate of the langage. When the set of minimal forbidden factors is infinite, as for ternary square-free words, Noonan and Zeilberger [6] propose to obtain an upper bound on the growth rate the as follows. Let L be a factorial langage with an infinite set S of minimal forbidden factors. We choose a finite subset S' of S, for example take S' as the set all of words in S of length at most ℓ, for some ℓ. Let L' be the factorial langage whose set of minimal forbidden factors is S'. We have $L \subset L'$, so we can compute the growth rate of L', which is also an upper bound on the growth rate of L.

Richard and Grimm [5] have extended the computations in [6], they obtained that the growth rate of ternary words avoiding squares xx with $|x| \leq 24$ is 1.30193812..

In Section 2, we describe a new method to obtain upper bounds on growth rates. In Section 3, we discuss the computation of an upper bound of 1.30178858.. on the growth rate of ternary square-free words.

2 Method for upper bounds

Let L be a factorial langage over Σ_s and let t_n be the number of words of length n in L. Let $k \geq 1$ and $\ell \geq 1$ be integers. We consider the t_k factors w_i in L of length k in lexicographic order. We define the $t_k \times t_k$ matrix $M_{k,\ell}$, as follows. For $0 \leq i < k$, $0 \leq j < k$, $M_{k,\ell}(i,j)$ is the number of words in L of length $k+\ell$ having w_i as a prefix and w_j as a suffix. For $n \geq 0$, we also define V_k^n as the vector of dimension t_k such that $V_k^n(i)$ is the number of words in L of length n having w_i as a suffix, for $0 \leq i < k$.

So $V_k^n = 0$ for $n < k$ and $V_k^k = {}^t[1,1,\ldots,1]$. Moreover, for $n \geq k$, we have $V_k^{n+\ell} \leq M_{k,\ell} \cdot V_k^n$, where the inequality is componentwise. To see this, notice that $M_{k,\ell}$ acts like adding a factors of length ℓ to the words in L. By definition of $M_{k,\ell}$, $V_k^{k+\ell} = M_{k,\ell} \cdot V_k^k$, but the inequality can be strict for $n > k$ because e.g. the minimal forbidden factors of length greater than $k+l$ are not taken into account.

Iterating q times this inequality yields to $V_k^{k+q\ell} \leq M_{k,\ell}^q \cdot V_k^k$. Let Ek,ℓ denote the largest eigenvalue of $M_{k,\ell}$. Let $|.|$ be the norm obtain by summing up the components. We have $t_n = |V_k^n|$, so that

$$
\begin{aligned}
\lim_{n\to\infty} (t_n)^{\frac{1}{n}} &= \lim_{q\to\infty} |V_k^{k+q\ell}|^{\frac{1}{k+q\ell}} \\
&\leq \lim_{q\to\infty} |M_{k,\ell}^q \cdot V_k^k|^{\frac{1}{k+q\ell}} \\
&\leq \lim_{q\to\infty} \left(kE_{k,\ell}^q\right)^{\frac{1}{k+q\ell}} \\
&= \lim_{q\to\infty} E_{k,\ell}^{\frac{q}{k+q\ell}} \\
&= E_{k,\ell}^{\frac{1}{\ell}}.
\end{aligned}
$$

To obtain upper bounds on the growth rate of L, we proceed as follows:

1. Choose integers k and ℓ.
2. Construct the matrix $M_{k,\ell}$ using a depth-first traversal of the set of words in L of length at most k.
3. Compute $E_{k,\ell}$ using the well-known power method, and finally compute $E_{k,\ell}^{\frac{1}{\ell}}$.

3 Application to repetition-free words

We associate to a ternary square-free word t of length $n+2$ a binary word b of length n defined as follows: for $0 \leq i < n$, $b[i] = 0$ if $t[i] = t[i+2]$ and $b[i] = 1$ otherwise. See the example below.

t	0	1	2	0	2	1	0	1	2	1	0	2
b	1	1	0	1	1	0	1	0	1	1		

This encoding is used in [3, 4] to prove parts of Dejean's conjecture. Let t_n (resp. t'_n) denote the number of ternary square-free words (resp. codes of ternary square-free words) of length n. Two ternary square-free words have the same code if and only if they are equal up to a permutation of the letters in Σ_3, so we have $t_{n+2} = 6t'_n$. This implies that the langage of ternary square-free words and the langage of their code have the same growth rate:

$$\begin{aligned}
\lim_{n\to\infty} (t_n)^{\frac{1}{n}} &= \lim_{n\to\infty} (t_{n+2})^{\frac{1}{n+2}} \\
&= \lim_{n\to\infty} \left(6t'_n\right)^{\frac{1}{n+2}} \\
&= \lim_{n\to\infty} \left(t'_n\right)^{\frac{1}{n}} .
\end{aligned}$$

Considering the codes instead of the ternary square-free words reduces the size of the matrix $M_{k,\ell}$ within a factor $6^2 = 36$ for the same parameters k and ℓ.

The size of the matrix is indeed the limiting factor here: the quality of the upper bound obtained mostly depends on k and the dimension of the matrix is exponential in k. The following observation is useful: If $\ell \leq k$, then every $M_{k,\ell}(i,j)$ is at most 1, since there exist at most one word of length $k+\ell$ with fixed prefix and suffix of size k. For codes of ternary square-free words, it remains true even for $\ell \leq k+1$. This allows to use only one bit per coefficient.

For $k = 44$ and $\ell = 45$, the matrix $M_{44,45}$ occupies 360979^2 bits, which is about 15.17 giga-bytes. The computation [1] took nearly three days on a 4-processor computer with 16 giga-bytes of memory. The construction of the matrix was parallelized.
We obtain an upper bound on the growth rate of ternary square-free words of 1.30178858.. which improves upon the previously best known of 1.30193812.. [5]. It is noticeable that Richard and Grimm [5] get an estimation of this growth rate of 1.301762.. using differential approximants on the known values of t_n for $n \leq 110$. Moreover, Kolpakov recently proved a lower bound of 1.30125 [2].

We also considered (codes of) binary cube-free words. For $k = \ell = 32$, we obtained an upper bound on the growth rate of binary cube-free words of 1.45758131.. which has to be compared to Kolpakov's lower bound of 1.456975 [2].

[1] The program is written in C with pthread and GMP. It is available at: http://dept-info.labri.fr/~ochem/morphisms/

References

[1] J. Berstel. Growth of repetition-free words - a review. Number 20 in Publications du Laboratoire de Combinatoire et d'Informatique Mathématique. Université du Québec à Montréal, February 1995.

[2] R. Kolpakov. On the number of repetition-free words, Workshop on Words and Automata, St Petersburg, June 7th, 2006.

[3] J. Moulin-Ollagnier. Proof of Dejean's conjecture for alphabets with $5, 6, 7, 8, 9, 10$ and 11 letters, *Theoret. Comput. Sci.* **95** (1992), 187–205.

[4] J.-J. Pansiot. A propos d'une conjecture de F. Dejean sur les répétitions dans les mots, *Disc. Appl. Math.* **7** (1984), 297–311.

[5] C. Richard and U. Grimm. On the entropy and letter frequencies of ternary square-free words, *Electron. J. Comb.* **11** (2004), #R14

[6] J. Noonan and D. Zeilberger. The Goulden-Jackson Cluster Method: Extensions, Applications, and Implementations, *J. Difference Eq. Appl.* **5** (1999), 355-377.