Journal of Integer Sequences, Vol. 5 (2002),
Article 02.2.2

The minimal density of a letter in an infinite ternary square-free word is $0.2746\cdots$

Yuriy Tarannikov

Mech. & Math. Department
Moscow State University
119992 Moscow, Russia

yutaran@mech.math.msu.su
taran@vertex.inria.msu.ru

Abstract

We study the minimal density of letters in infinite square-free words. First, we give some definitions of minimal density in infinite words and prove their equivalence. Further, we propose a method that allows to strongly reduce an exhaustive search for obtaining lower bounds for minimal density. Next, we develop a technique for constructing square-free morphisms with extremely small density for one letter that gives upper bounds on the minimal density. As an application of our technique we prove that the minimal density of any letter in infinite ternary square-free words is $0.2746\cdots$.

A word is called *square-free* if it cannot be written in the form $axxb$ for two words a, b and a nonempty word x. It is easy to see that the maximal length of a binary square-free word is 3. A. Thue proved [8] that there exist ternary square-free words. The number of ternary square-free words of length n is given by the sequence **A006156** in The Encyclopedia of Integer Sequences [7]. Ekhad and Zeilberger [2] proved that the number of ternary square-free words of length n is at least $2^{n/17}$. Grimm [3] gave a better bound; he proved that this number is at least $65^{n/40}$. Note that not every finite square-free word can be extended to an infinite square-free word. In this paper we prove that the minimal density of any letter in an infinite ternary square-free word is $0.2746\cdots$.

1 Preliminary concepts and notions

Let M be a finite alphabet, and let M^* be the free monoid over M. Let M^ω be the set of one-sided infinite words over M (or mappings from $\mathbf{N} \to M$). A word $v \in M^*$ is called a *factor* of the word $w \in M^*$ if w can be written as $w = v_1 v v_2$ for some $v_1, v_2 \in M^*$. If v_1 is the empty word, then v is called also a *prefix* of w. Let $F \subseteq M^*$. Denote by $l(F)$ the length of the longest word in F; if the set F is infinite we define $l(F) := \infty$. The set F for $F \subseteq M^*$ is called a *factorial language* if for any word $w \in F$ the set F contains every factor of w. The length of the word w is denoted by $|w|$.

Any set G of forbidden factors in M^* generates a factorial language $F = F(G)$ where

$$F(G) = \{w \in M^* \mid \forall v \in G \quad v \text{ is not a factor of } w\}.$$

Indeed, if the word $w \in F(G)$ does not contain any word $v \in G$ then no factor u of w contains such a word, either.

We denote by $F^\omega \subseteq M^\omega$ the set of all infinite words with every finite factor belonging to F.

Proposition 1.1 *The set F^ω is not empty iff the language F is infinite.*

Proof. If the language F is finite then obviously the set F^ω is empty. If the language F is infinite then we can construct an infinite word with every finite prefix belonging to F by König's Infinity Lemma. It is easy to see that this word belongs to F^ω. ∎

Denote by $a(w)$ the number of occurrences of a letter $a \in M$ in the word $w \in F$. The proportion $\rho_a(w) = \frac{a(w)}{|w|}$ is called the *density* of the letter a in the finite word w. To define the density of a letter in the infinite word w is a more complicated problem. We can consider the sequence $(\rho_a(w_n))$, $n = 1, 2, \ldots$, where w_n is the prefix of w of length n but it is possible that this sequence does not converge to a limit. An example of such a situation is given by the infinite word

$$w = a\underbrace{b\ldots b}_{10}\ \underbrace{a\ldots a}_{10^2}\ \underbrace{b\ldots b}_{10^4}\ \underbrace{a\ldots a}_{10^8}\ \underbrace{b\ldots b}_{10^{16}}\ \cdots$$

It is not hard to see that for any positive integer N, positive real ε and real ξ, $0 \le \xi \le 1$, there exists $n > N$ such that $|\rho_a(w_n) - \xi| < \varepsilon$.

Thus, it is possible that the limit of the densities for the sequence of prefixes in an infinite word does not exist. Nevertheless, we can define the lower limit of this sequence.

Definition 1.1 *Let F be an infinite factorial language. We define*

$$F(l) := \{w \in F \mid\ |w| = l\};$$

$$\rho_a(F, l) := \min_{w \in F(l)} \rho_a(w); \text{ and}$$

$$\rho_a(F) := \varliminf_{l \to \infty} \rho_a(F, l).$$

The next two lemmas are proved in Kolpakov, Kucherov, and Tarannikov [4].

Lemma 1.1 *For every $l \in \mathbf{N}$, the inequality $\rho_a(F,l) \le \rho_a(F)$ holds.*

Lemma 1.2 $\rho_a(F) = \lim_{l\to\infty} \rho_a(F,l) = \sup_{l\ge 1} \rho_a(F,l)$.

Thus, we can write

$$\rho_a(F) = \lim_{l\to\infty} \rho_a(F,l).$$

Definition 1.2 *We denote*

$$\mathcal{A}_a^-(\xi) = \{w \in F \,|\, \forall \text{ prefixes } u \text{ of } w:\ \rho_a(u) \le \xi\},$$

$$\mathcal{A}_a^-(\xi-) = \{w \in F \,|\, \forall \text{ prefixes } u \text{ of } w:\ \rho_a(u) < \xi\}.$$

Theorem 1.1 *If $\rho_a(F) \le \xi$ then the set $\mathcal{A}_a^-(\xi)$ is infinite.*

Proof. Assume the converse. Then we can decompose an arbitrary infinite word w in F^ω into $w = v_1 v_2 \dots$ where $|v_i| \le l(\mathcal{A}_a^-(\xi)) + 1$ and $\rho_a(v_i) \ge \rho_a(F) + \varepsilon$ for some $\varepsilon > 0$. Therefore,

$$\lim_{l\to\infty} \rho_a(F,l) \ge \rho_a(F) + \varepsilon.$$

This contradiction proves the theorem. ■

Corollary 1.1

(a) *The set $\mathcal{A}_a^-(\xi)$ is infinite iff $\xi \ge \rho_a(F)$.*

(b) *The set $\mathcal{A}_a^-(\xi-)$ is infinite iff $\xi > \rho_a(F)$.*

Corollary 1.2 *There exists a word $w \in F^\omega$ such that any prefix u of w belongs to $\mathcal{A}_a^-(\rho_a(F))$.*

Proof. We can construct easily this word w by König's Lemma on an infinite tree. ■

The above facts allow us to obtain lower bounds ξ for the minimal density of a letter a in a factorial language F proving by an exhaustive search that the set $\mathcal{A}_a^-(\xi)$ ($\mathcal{A}_a^-(\xi-)$) is finite. As it will be shown below in many cases for sufficiently small ξ this exhaustive search can be produced in a very short time.

2 Minimal letter density for some special factorial languages

For some special factorial languages the problem of finding the minimal letter density is (almost) trivial.

Example 2.1 The factorial language $F = F(G)$ is generated by a finite set G of prohibited factors. Then the minimal letter density $\rho_a(F)$ is rational and equal to the minimal density of the letter a over all cycles (accessible from the starting vertex) in the transition graph of language $F(G)$. (For transition graphs of factorial languages see Rosaz [6].)

Example 2.2 Let $M = \{0, 1\}$, and let ξ be a real number with $\xi \in (0; 1)$. Let F be the set of all finite factors of a standard Sturmian word $(a_1 a_2 \cdots)$ where $a_i = \lfloor (i+1)\xi \rfloor - \lfloor i\xi \rfloor$, $i = 1, 2, \ldots$. Then any infinite word in F^ω has density ξ. So, $\rho_1(F) = \xi$.

Example 2.3 Let $M = \{a, b\}$ and let F be the set of all overlap-free binary words (i. e., words that do not contain a factor w that has the form $w = v_1 v_2 c$ where c is the first letter of the word v_1). Restivo and Salemi [5] proved that any infinite overlap-free binary word is a concatenation of factors (ab) and (ba) with a preperiod of one or two symbols. It follows that $\rho_a(F) = \rho_b(F) = 1/2$.

Example 2.4 Infinite square-free words on the alphabet M (i. e., words that do not contain a factor w that has the form $w = vv$). It is obvious that if $|M| = 2$ then there are no infinite square-free words over alphabet M. There exist infinite square-free words on the ternary alphabet. A. Thue was the first to construct an example of such a word [8]. Therefore, if $|M| \geq 4$ we can construct an infinite square-free word over the alphabet $M \setminus \{a\}$. Consequently, $\rho_a(F) = 0$. Thus, the only interesting case in this respect is $|M| = 3$.

3 Lower bounds for the minimal letter density in ternary square-free words

In what follows F denotes the set of all ternary square-free words. The technique used to obtain the results given in this section was developed in Section 1. In the following table we give calculated values of numbers $l(\mathcal{A}_a^-(\xi-))$ and $l(\mathcal{A}_a^-(\xi))$ for "critical" ξ (i. e., for ξ such that these numbers differ). In our computer search we used the standard backtracking technique. For $\xi > 39/142$ we did not calculate all "critical" values of ξ because of the increasing of the required computer time.

Theorem 3.1 $l(\mathcal{A}_a^-(1780/6481-)) = 17312$.

The last result took near 40 hours on a Pentium, 166 MHz.

Corollary 3.1 *Let F be the set of ternary square-free words. Then $\rho_a(F) \geq 1780/6481 = 0.274648\cdots$ for all letters a.*

ξ (Proportion)	ξ (Decimal)	$l(\mathcal{A}_a^-(\xi-))$	$l(\mathcal{A}_a^-(\xi))$
0	0	0	3
1/4	0.25	3	15
4/15	0.266666	15	59
16/59	0.271186	59	63
3/11	0.272727	63	74
20/73	0.273973	74	136
37/135	0.274074	136	198
54/197	0.274112	198	252
17/62	0.274194	252	307
14/51	0.274510	307	324
81/295	0.274576	324	771
67/244	0.274590	771	801
53/193	0.274611	801	1034
145/528	0.274621	1034	1318
92/335	0.274627	1318	1481
407/1482	0.274629	1481	1500
354/1289	0.274631	1500	1765
485/1766	0.274632	1765	1784
170/619	0.274637	1784	2028
549/1999	0.274637	2028	2494
209/761	0.274639	2494	2778
613/2232	0.274642	2778	3488
691/2516	0.274642	3488	3772
443/1613	0.274644	3772	4168
950/3459	0.274646	4168	4715
1028/3743	0.274646	4715	4999
1223/4453	0.274646	4999	5709
1301/4737	0.274646	5709	5993
1496/5447	0.274647	5993	6703
1574/5731	0.274647	6703	6987
1769/6441	0.274647	6987	7383
1847/6725	0.274647	7383	7667
39/142	0.274647	7667	10882
1780/6481	0.274648	17312	

Table 1. Numbers $l(\mathcal{A}_a^-(\xi-))$ and $l(\mathcal{A}_a^-(\xi))$ for some “critical” ξ.

4 Upper bound

The most natural way to prove an upper bound for the minimal letter density is to construct a concrete word. As a result the letter density in this word will be a desired upper bound.

One of the main ways for constructing concrete infinite words is to use expansive morphisms. We consider morphisms of the form $h: M^* \to M^*$. In our case $M = \{a, b, c\}$. For $d \in M$ the infinite word $h^*(d)$ is generated by the infinite sequence of its prefixes

$$d, h(d), h(h(d)), h(h(h(d))), \ldots .$$

A morphism h is called *square-free* if $h(w)$ is square-free whenever w is square-free. If h is a square-free morphism then, obviously, for any letter $d \in M$ the word $h^*(d)$ will be square-free. If for any letter $m \in M$ we have $\rho_a(h(m)) = \xi$ then, obviously, $\rho_a(h^*(d)) = \xi$ too. The words $h(m)$ are finite, therefore here $\xi = \frac{p}{q}$ is rational. Thus, the simplest way is to try to construct a morphism where images of all letters consist of fragments of lengths q that contain the letter a exactly p times for some positive integers p and q. The problem is how to choose p and q? Here we give an empirical method of selecting good parameters p and q.

Let $\xi = \frac{p}{q}$ be a rational number. Define

$$\mathcal{A}_a^-(\xi*) = \{w \in F \mid \forall \text{ prefixes } u \text{ of } w: \ \rho_a(u) < \xi \text{ and if } |u| = nq, n \in \mathbf{N}, \text{ then } \rho_a(u) = \xi\}.$$

For a given rational ξ we search an (infinite) word in the set $\mathcal{A}_a^-(\xi*)$ by the usual backtracking technique. In many cases we obtain in a short time that the set $\mathcal{A}_a^-(\xi*)$ is finite. Thus, we cannot apply the proposed method for the construction of a morphism. If the length of the maximal found prefix increases with stable high speed then we do an empirical conclusion that a morphism with proportion p to q can exist. If the length of the maximal found prefix increases very slowly we conclude that probably such a morphism does not exist.

This empirical method can be applied to the problem of minimal letter density in any factorial language. At first, we tried it for ternary square-free words. We obtained very strong empirical confirmation for the ratio 64/233. The set $\mathcal{A}_a^-(64/233*)$ contains only 10 words of length 233 (5 up to replacing $b \leftrightarrow c$). Combining these words we tried to construct a square-free morphism. We used the following test of Bean, Ehrenfeucht, and McNulty [1] that guarantees that a morphism is square-free:

> **<u>If</u>**
>
> (0) $h(w)$ is square-free whenever w is a word on M which is square-free and of length not greater than three,
>
> and
>
> (1) $a = b$ whenever $a, b \in M$ with $h(a)$ a subword of $h(b)$
>
> **<u>then</u>**
>
> $h(u)$ is square-free whenever u is a square-free word on M.

In our case it is sufficient to check that
(1) $h(a)$, $h(b)$ and $h(c)$ are not factors of one another,
(0) the words

$$\begin{array}{lll} h(a)h(b)h(a) & h(b)h(a)h(b) & h(c)h(a)h(b) \\ h(a)h(b)h(c) & h(b)h(a)h(c) & h(c)h(a)h(c) \\ h(a)h(c)h(a) & h(b)h(c)h(a) & h(c)h(b)h(a) \\ h(a)h(c)h(b) & h(b)h(c)h(b) & h(c)h(b)h(c) \end{array}$$

are square-free.

The desired morphism was constructed. We give it here. Denote

$A = bcbacbcabcbabcacbcabcbacbcacbabcbacbcabcbabcacbabcbacbcacbabcacbcabcbacbcacb$
$abcbacbcabcbabcacbcabcbacbcacbabcacbcabcbabcacbabcbacbcabcbabcacbcabcbacbcacbabc$
$bacbcabcbabcacbabcbacbcacbabcacbcabcbacbcacbabcbacbcabcbabcacbcabcbacbcabacba$

$B = bcbacbcabcbabcacbcabcbacbcacbabcbacbcabcbabcacbabcbacbcacbabcacbcabcbacbcacb$
$abcbacbcabcbabcacbcabcbacbcacbabcacbcabcbabcacbabcbacbcabcbabcacbcabcbacbcacbabc$
$bacbcabcbabcacbabcbacbcacbabcacbcabcbacbcacbabcbacbcabcbabcacbcabcbacbcacbaca$

$B' = cbcabcbacbcacbabcbacbcabcbabcacbcabcbacbcacbabcacbcabcbabcacbabcbacbcabcbabc$
$acbcabcbacbcacbabcbacbcabcbabcacbabcbacbcacbabcacbcabcbacbcacbabcbacbcabcbabcacb$
$cabcbacbcacbabcacbcabcbabcacbabcbacbcabcbabcacbcabcbacbcacbabcbacbcabcbabcaba$

$C = bcbacbcabcbabcacbcabcbacbcacbabcbacbcabcbabcacbcabcbacbcabacbabcbacbcabcbabc$
$acbcabcbacbcacbabcbacbcabcbabcacbabcbacbcacbabcacbcabcbacbcacbabcbacbcabcbabcacb$
$cabcbacbcacbabcacbcabcbabcacbabcbacbcabcbabcacbcabcbacbcacbabcbacbcabcbabcaba$

$D = bcbacbcabcbabcacbcabcbacbcacbabcbacbcabcbabcacbcabcbacbcabacbabcbacbcabcbabc$
$acbcabcbacbcacbabcbacbcabcbabcacbabcbacbcacbabcacbcabcbacbcacbabcbacbcabcbabcacb$
$cabcbacbcacbabcacbcabcbabcacbabcbacbcabcbabcacbcabcbacbcacbabcbacbcabcbacabca$

$D' = cbcabcbacbcacbabcbacbcabcbabcacbcabcbacbcacbabcbacbcabcbacabcacbcabcbacbcacb$
$abcbacbcabcbabcacbcabcbacbcacbabcacbcabcbabcacbabcbacbcabcbabcacbcabcbacbcacbabc$
$bacbcabcbabcacbabcbacbcacbabcacbcabcbacbcacbabcbacbcabcbabcacbcabcbacbcabacba$

(the words B' and D' can be derived from B and D respectively by replacing $b \leftrightarrow c$).

The constructed morphism h is

$$\begin{aligned} h(a) &= BA \\ h(b) &= BDB'D' \\ h(c) &= BCB'D' \end{aligned}$$

As a result of this construction we conclude that

Theorem 4.1 *Let F be the set of ternary square-free words. Then $\rho_a(F) \leq 64/233 = 0.274678\cdots$ for all letters a.*

In combination with the lower bound given in Corollary 4.1 we have

Theorem 4.2 *Let F be the set of ternary square-free words. Then $0.274648\cdots = 1780/6481 \leq \rho_a(F) \leq 64/233 = 0.274678\cdots$ for all letters a.*

Thus, $\rho_a(F) = 0.2746\cdots$.

5 Acknowledgments

The author is grateful to the anonymous referee for the detailed comments which improved the readability of the paper.

References

[1] D. R. Bean, A. Ehrenfeucht, and G. F. McNulty, Avoidable patterns in strings of symbols, *Pacific J. Math.* **85** (1979), 261–294.

[2] S. B. Ekhad and D. Zeilberger, There are more than $2^{n/17}$ n-letter ternary square-free words, *J. Integer Sequences* **1** (1998), Article 98.1.9.

[3] U. Grimm, Improved bounds on the number of ternary square-free words, *J. Integer Sequences* **4** (2001), Article 01.2.7.

[4] R. Kolpakov, G. Kucherov, and Yu. Tarannikov, On repetition-free binary words of minimal density, *Theoret. Comput. Sci.* **218** (1999), 161–175.

[5] A. Restivo, S. Salemi, On weakly square free words, *Bulletin of the EATCS*, No. 21 (1983), 49–56.

[6] L. Rosaz, Unavoidable sets of words, Thèse de doctorat, Université Paris 7, 1993.

[7] N. J. A. Sloane, S. Plouffe, *The Encyclopedia of Integer Sequences*, Academic Press, 1995. See also On-line Encyclopedia of Integer Sequences. Available at `http://www.research.att.com/~njas/sequences/`.

[8] A. Thue, Über unendliche Zeichenreihen, *Norske Videnskabers Selskabs Skrifter I, Matematisk-Naturvidenskapelig Klasse*, Kristiania, **7** (1906), 1–22.

2000 *Mathematics Subject Classification*: 11B05 .
Keywords: Combinatorics on words; square-free word; factorial languages; minimal density

---

(Concerned with sequence A006156.)

Received March 21, 2002; revised version received October 20, 2002. Published in Journal of Integer Sequences October 24, 2002.

Return to Journal of Integer Sequences home page.