Theoretical Computer Science 23 (1983) 69–82
North-Holland Publishing Company

UNIFORMLY GROWING k-TH POWER-FREE HOMOMORPHISMS

Franz-Josef BRANDENBURG

Institut für Informatik, Universität Bonn, 5300 Bonn, Fed. Rep. Germany

Communicated by A. Salomaa
Received July 1981
Revised February 1982

Abstract. A string is called kth power-free, if it does not have x^k as a nonempty substring. For all nonnegative rational numbers k, kth power-free strings and kth power-free homomorphisms are investigated and the shortest uniformly growing square-free ($k = 2$) and cube-free ($k = 3$) homomorphisms mapping into least alphabets with three and two letters are introduced. It is shown that there exist exponentially many square-free and cube-free strings of each length over these alphabets. Sharpening the kth power-freeness to the repetitive threshold $\mathrm{RT}(n)$ of n letter alphabets, we provide arguments for the nonexistence of various $\mathrm{RT}(n)$th power-free homomorphisms.

1. Introduction

Since the work of Thue [9, 10] at the beginning of this century there have been many investigations on the construction of strings without repetitions. The simplest such strings are the square-free and the cube-free strings, which do not have x^2 and x^3 as a nonempty substring. Curiously the English words square-free and repetitive each have a repetition and they are examples of non square-free strings, such as the mathematical constants $e = 2.7182\,8182\,8\ldots$, $\sqrt{3} = 1.7320\,5080\,8\ldots$, and $\pi = 3.1415\,9265\,3589\,7932\,3846\,2643\,3\ldots$ (see [7]). The existence of square-free strings of arbitrary length over three letter alphabets and of cube-free strings over two letter alphabets has originally been discovered by Thue [9, 10]. This is in fact surprising and a remarkable combinatorial property of strings. Square-free strings have been applied in various situations, e.g., in unending chess, in group theory, and in formal language theory. Details can be found in [8] and the references given there.

In this paper we generalize the notion of a power of a string to rational powers and strict rational powers. Square-free and cube-free strings now are special cases with $k = 2$ and $k = 3$, and strongly cube-free or overlap-free strings are weakly 2nd power-free strings. This generalization is a real simplification over existing notions. Furthermore, it allows to define partial repetitions of strings and repetitive thresholds of alphabets.

0304-3975/83/0000–0000/$03.00

Particular emphasis is laid upon the sets of square-free and cube-free strings over least alphabets. It is easy to see that there are no square-free strings of length four over a two letter alphabet and no cube-free strings of length four over a single letter alphabet. However, there are infinite such strings over alphabets with three resp. two letters. Here we improve these results and show that for every positive integer n there are exponentially many square-free strings of length n over a three letter alphabet and exponentially many cube-free strings of length n over a two letter alphabet. Hence, the sets of square-free and cube-free strings are either trivial or exponentially dense.

For the proofs of these results we use square-free and cube-free homomorphisms from an arbitrary alphabet into the sets of strings over a three and a two letter alphabet, respectively. We introduce the shortest uniformly growing square-free homomorphisms from alphabets up to six letters into the set of strings over a three letter alphabet and the shortest cube-free homomorphisms from alphabets up to four letters into the set of strings over a two letter alphabet. Much longer and not uniformly growing homomorphisms of this kind have been introduced in [1]. We also sharpen conditions on uniformly growing homomorphisms to be square-free to the very optimum. This improves results by Bean *et al.* [1], Berstel [2] and Thue [9, 10].

Finally, the repetitive threshold $\mathrm{RT}(n)$ of n letter alphabets Σ_n is defined. $\mathrm{RT}(n)$ is the least k such that there are infinitely many weakly kth power-free strings over Σ_n. This continues the work of Déjean [3]. We establish certain properties of weakly $\mathrm{RT}(n)$th power-free homomorphisms. In particular, we show the non-existence of nontrivial such homomorphisms from Σ_{n+1}^* into Σ_n^*, and from Σ_n^* into Σ_n^*, if $\mathrm{RT}(n) < \frac{3}{2}$. These results imply that new techniques are necessary to determine the unknown values of the repetitive threshold $\mathrm{RT}(n)$ for $n \geq 4$, and they show that our proof techniques for establishing lower bounds on the numbers of kth power-free strings fail here.

2. kth power-free strings

For all nonnegative rational numbers k we define kth power-free strings and kth power-free homomorphisms and establish some basic properties.

For a string w over an alphabet Σ let $|w|$ denote the *length* of w. A string x is a prefix (*substring*, *suffix*) of w, if $w = xy$ ($w = uxv$, $w = yx$) for strings u, v and y. For an integer n define the nth *power* of w by $w^0 = \lambda$ and $w^n = ww^{n-1}$, where λ denotes the *empty string*. This notion is generalized to rational exponents.

Definition. Let w be a string over Σ. Let k be a nonnegative rational number and let $n \geq k$ be an integer. Then the kth *power* (*strict kth power*) of w is the least prefix x of w^n such that $|x| \geq k \cdot |w|$ ($|x| > k \cdot |w|$).

For example, *aba* is the $\frac{3}{2}$th power of *ab*, and *abab* is the strict $\frac{3}{2}$th power of *ab*. If w is a nonempty string and a is the first letter of w, then wa is the strict first power of w.

In the following we shall see that the use of rational and strict rational powers of strings both generalizes and simplifies existing notions.

Definition. A (finite or infinite) string w is *kth power-free* (*weakly kth power-free*), if w does not have the (strict) kth power of a non-empty string as a substring, i.e., $w \neq ux^k v$, where k is a nonnegative rational number. In accordance with the commonly used terminology, second and third power-free strings are called *square-free* and *cube-free*, respectively.

Let $\mathrm{FREE}_\Sigma(<k)$, $\mathrm{FREE}_\Sigma(\leqslant k)$ and $\mathrm{FREE}_\Sigma(=k)$ denote the *sets of kth power-free strings, weakly kth power-free strings* and *exactly kth power-free strings* over the alphabet Σ, respectively, where $\mathrm{FREE}_\Sigma(=k) = \mathrm{FREE}_\Sigma(\leqslant k) - \mathrm{FREE}_\Sigma(<k)$. When it is appropriate we replace the subscript Σ by its cardinality. Thus $\mathrm{FREE}_3(<2)$, e.g., denotes the set of square-free or second power-free strings over any fixed three letter alphabet.

The languages $\mathrm{FREE}_\Sigma(<k)$ and $\mathrm{FREE}_\Sigma(\leqslant k)$ consists of all strings over Σ, which have at most mth powers, where $m < k$ and $m \leqslant k$, respectively. Every string w in $\mathrm{FREE}_\Sigma(=k)$ has a kth power, i.e., $w = ux^k v$, but w does not have an mth power with $m > k$. Obviously, every kth power-free string is weakly kth power-free, and every weakly kth power-free string is mth power-free, if $k < m$. Thus $\mathrm{FREE}_\Sigma(<k) \subseteq \mathrm{FREE}_\Sigma(\leqslant k) \subseteq \mathrm{FREE}_\Sigma(<m)$.

It is easy to see that $\mathrm{FREE}_1(\leqslant k)$ contains only strings up to length $[k+1]$, and that $\mathrm{FREE}_2(2)$ contains seven strings up to length three. To the contrary, Thue [9] has discovered the existence of infinite square-free strings over three letter alphabets and of infinite cube-free strings over two letter alphabets.

One of the aims pursued in this paper is to count square-free strings over a three letter alphabet and cube-free strings over a two letter alphabet, i.e., to establish lower and upper bounds on the numbers of such strings of length n for every n. A trivial upper bound stems from the number of all strings, which grows exponentially in the cardinality of the alphabet. For our proofs of exponential lower bounds we make use of kth power-free homomorphisms, which are the most useful tools in the theory of kth power-free strings.

A *homomorphism* h is a mapping between free monoids Σ^* and Δ^* with $h(xy) = h(x)h(y)$ for every $x, y \in \Sigma^*$. A homomorphism h is *length uniform*, if $|h(a)| = |h(b)|$ for every $a, b \in \Sigma$. h is *growing*, if $h(a) \neq \lambda$ for every $a \in \Sigma$ and $|h(a)| > 1$ for some $a \in \Sigma$, and h is *uniformly growing*, if h is length uniform and growing, i.e., $|h(a)| = t > 1$ for every $a \in \Sigma$.

A homomorphism is compatible with the product of strings, but it is not compatible with rational powers of strings. Hence, it may occur that $h(w)^k \neq h(w^k)$. As an example consider $k = \frac{3}{2}$, $h(a) = a$, $h(b) = bcd$, $x = ab$ and $y = ba$. Then $h(x^k) =$

$h(aba) = abcda$, $h(x)^k = abcd^k = abcdab$, $h(y^k) = h(bab) = bcdabcd$, and $h(y)^k = bcda^k = bcdabc$. Thus the repetitive power of a string may increase or decrease under a homomorphism.

Definition. A homomorphism h is *(weakly) kth power-free*, if $h(\mathrm{FREE}_\Sigma(<k)) \subseteq \mathrm{FREE}_\Delta(<k)$ $(h(\mathrm{FREE}_\Sigma(\leq k)) \subseteq \mathrm{FREE}_\Delta(\leq k))$.

Every (weakly) kth power-free homomorphism maps (weakly) kth power-free strings into such strings. It may however occur that a (weakly) kth power-free homomorphism is not (weakly) mth power-free, where $m > k$ or $m < k$. Examples are easy to find due to the incompatibility mentioned above. However, (weakly) kth power-free homomorphisms can be composed without harm.

Theorem 1. *If h_1 and h_2 are (weakly) kth power-free homomorphisms, then so is their composition $h_1 \cdot h_2$.*

This property provides us with a powerful and elegant tool for defining infinite (weakly) kth power-free strings, and we (as others before) make use thereof. In particular, if h is growing and $h(a) = ax$ for a letter a and a nonempty string x, then the limit of the sequence $h^n(a) = h(h^{n-1}(a))$ is an infinite (weakly)kth power-free string, if h is such a homomorphism. Furthermore, Thue [10], Bean *et al.* [1] and Berstel [2] have established conditions which guarantee that a homomorphism is kth power-free. These conditions say that it is sufficient to check the kth power-freeness of a homomorphism on all kth power-free strings of length $k+1$ and either a certain substring property of the homomorphic images of the letters (see [1, 10]) or, for square-free homomorphisms, the square-freeness on all square-free strings of length $2 + 2p(h)$, where $p(h) = \max\{h(a) \mid a \in \Sigma\}/\min\{h(a) \mid a \in \Sigma\}$ (see [2]). These conditions are too restrictive for uniformly growing square-free homomorphisms and optimal ones are established by Theorem 2.

Theorem 2. *Let h be a uniformly growing homomorphism with domain Σ^*. Then h is square-free if and only if $h(w)$ is square-free for every square-free string w of length three.*

Proof. The proof is given for arbitrary homomorphisms and we shall point out, where the length uniformity is needed and where it simplifies the proof.

Let h be square-free on every square-free string w with $|w| = 3$. Then $h(a) \neq \lambda$ for every letter a, and $h(a)$ is not a prefix (suffix) of $h(b)$ for every $a, b \in \Sigma$ with $a \neq b$; otherwise, $h(bab)$ has a square.

Suppose that $h(w)$ has a square and that $h(w)$ has minimal length. Then $h(w) = xttz'$. By the minimality, $xx' = h(a)$ and $zz' = h(c)$ for some letters a and c, and by the hypothesis $|w| \geq 4$, so that $w = aw'c$. From the length-uniformity of h we obtain that $w' = ubv$ with $b \in \Sigma$, $h(b) = yy'$, and $x'h(u)y = y'h(v)z = t$. For arbitrary homomorphsims w' does not necessarily have this form. Note that the length-uniformity of h will not be used elsewhere.

If $|x'|=|y'|$, then $x'=y'$, $h(u)=h(v)$, $u=v$ and $y=z$. Now $a=b$ or $b=c$, since $a\neq b\neq c$ implies $h(abc)=xx'yx'yz'$. Hence $w=bubuc$ or $w=aubub$, and w has a square. Let $|x'|\neq|y'|$ and assume $|x'|>|y'|$. The case $|x'|<|y'|$ is similar. Then $x'=y'x''$ with $x''\neq\lambda$, and $v=v'dv''$ with $d\in\Sigma$, $|h(v')|<|x''|$ and $|h(v'd)|\geq|x''|$. If h is length uniform, then $v'=\lambda$, which simplifies this analysis a bit. Now $h(d)=\delta\delta'$ and $h(a)=xy'h(v')\delta$, where $|\delta|=|x''|-|h(v')|>0$. Thus $h(ad)$ has a square, which implies that $a=d$. Hence, $h(a)$ begins and ends with δ and $2\cdot|\delta|<|h(a)|$, i.e., $h(a)=\delta\beta\delta$ with $\beta\neq\lambda$; otherwise, $h(a)$ has a square by results in [5, 6]. Furthermore, β is a prefix of $h(ub)$. Thus $h(aub)=\delta\beta\delta\beta\gamma$ has a square, contradicting the minimality of $h(w)$. Hence, h is square-free.

It is easy to see that the homomorphism g with $g(a)=ab$, $g(b)=cb$ and $g(c)=cd$ is square-free on all square-free strings of length one or two, but g is not square-free on abc. Thus the bound three is optimal. □

For non uniformly growing homomorphisms consider the homomorphism g from Example 1.6 in [1], which maps a, b, c, d, e to ad, b, $cdbadce$, $cdabdce$, $cdadbce$, respectively. g is square-free for all square-free strings of length three. However, $g(abac)=adbadcdbadce$ has the square $(dbadc)^2$. Here, $w=abac$ factors into auc with $g(a)=xx'$ and $g(c)=yx'h(u)yz$. Thus $g(c)$ forces the square and prevents the factorization of w into $aubvc$ with the properties as in Theorem 2.

Note that the substring property of Thue and Bean et al. [10, 1] is too restrictive. This property guarantees the square-freeness of every homomorphism. It requires that $h(a)$ is not a substring of $h(b)$ for letters a, b with $a\neq b$, which means that the homomorphic image of each letter can uniquely be determined in a string. However, the homomorphism h with $h(a)=a$ and $h(b)=bac$, e.g., is kth power-free for every $k>\frac{3}{2}$, and h violates the substring property from above, as does the 'Thue' homomorphism, which is of interest for its own right.

Theorem 3. *The 'Thue' homomorphism h_1 with $h_1(a)=ab$ and $h_1(b)=ba$ is the shortest uniformly growing (weakly) kth power-free homomorphism for every $k>2$ $(k\geq 2)$.*

Proof. Let $k=2+p+r$ with $0\leq r<1$. Suppose that $h_1(w)$ has the kth power of a nonempty string as a substring and is of minimal length. Then $h_1(w)=cv^{2+p}v^rd$ with $c, d\in\{a,b\}\cup\{\lambda\}$. If $|v|$ is odd, then there exists no x such that $h_1(x)=cvve$, where either $c=e=\lambda$ or $c, e\in\{a,b\}$ and e is the first letter of v. Hence, $|v|$ is even. Then $c=\lambda$ by the minimality of $h_1(w)$ and $w=x^{2+p}y$, where $x=h_1^{-1}(v)$ and $y=h_1^{-1}(v^r)$, if $|v^r|$ is even, and $y=h_1^{-1}(v^rd)$, if $|v^r|$ is odd. Then $y=x^r$ and $w=x^{2+p}y$ has a kth power, which is a strict kth power, if $h_1(w)$ has a strict kth power. Hence, h_1 is (weakly) kth power-free. □

We close this section with examples of two particular weakly kth power-free homomorphisms, which are optimal in a sense made clear below.

Example 1. Let $h_1(a) = ab$ and $h_1(b) = ba$. h_1 is called the 'Thue' homomorphism, who has studied the string obtained by iterating h_1. See [8–10]. In particular, h_1 is weakly square-free and defines an infinite weakly square-free string over $\{a, b\}$ by iteration. Notice that there are no growing square-free homomorphisms mapping into the set of strings over two letter alphabets, since this set is finite.

Example 2. Let

$$h_2(a) = abcacbcabcbacbcacba,$$

$$h_2(b) = bcabacabcacbacabacb,$$

$$h_2(c) = cabcbabcabacbabcbac.$$

The homomorphism h_2 is due to Déjean [3]. She has shown that h_2 is weakly $\frac{7}{4}$th power-free, and that $\frac{7}{4}$th power-free strings over three letter alphabets are of length at most 38. Thus there exists an infinite (or infinitely many) weakly $\frac{7}{4}$th power-free string over a three letter alphabet, but there exist no growing $\frac{7}{4}$th power-free homomorphisms over three letter alphabets. Under various perspectives h_2 is an extension of h_1 to three letter alphabets; details are discussed in Section 4.

3. Density of square-free and cube-free sets of strings

Our first investigation and experience with square-free strings was the attempt of computing all initial square-free strings over $\{a, b, c\}$ and listing these strings as the paths of a tree with root λ. See [10]. This representation immediately asks for bounds on the width of the tree so obtained, which equals the number of square-free strings of length n.

Theorem 4. *For every alphabet Σ there exists a uniformly growing square-free homomorphism h from Σ^* into $\{a, b, c\}^*$.*

Proof. Let $\Sigma = \{a_1, a_2, \ldots, a_n\}$ and define homomorphisms $h_1, \ldots, h_6$ by the following table:

	h_1	h_2	h_3	h_5	h_6
a_1	ab	ab	$abacbabcbac$	$abacabcacbabcbacbc$	$abacabcacbabcbacabacbc$
a_2		ac	$abacbcacbac$	$abacabcacbacabacbc$	$abacabcacbcabcbabcacbc$
a_3			$abcbabcacbc$	$abacabcacbcabcbabc$	$abacabcbabcacbabcbacbc$
a_4				$abacabcbacabacbabc$	$abacabcbabcacbcabcbabc$
a_5				$abacabcbacbcacbabc$	$abacabcbacabacbabcacbc$
a_6					$abacabcbacbcacbabcacbc$

Each of these homomorphisms is square-free on all square-free strings of length three, and thus square-free by Theorem 2.

If Σ has more than six letters, then define copies $h_6^{(i)}$ of h_6 from $\{a_{6i-5}, \ldots, a_{6i}\}^*$ into $\{a_{3i-2}, a_{3i-1}, a_{3i}\}^*$, identifying a and a_{3i-2}, b and a_{3i-1}, and c and a_{3i} under $h_6^{(i)}$. Suppose that Σ has $3 \cdot 2^p$ letters and let $r = 2^{p-1}$. For homomorphisms g_1, g_2 from Σ_i^* into Δ_i^* with $\Sigma_1 \cap \Sigma_2 = \Delta_1 \cap \Delta_2 = \emptyset$ define the *parallel composition* $g_1 \times g_2$ from $(\Sigma_1 \cup \Sigma_2)^*$ into $(\Delta_1 \cup \Delta_2)^*$ by $(g_1 \times g_2)(a) = g_1(a)$, if $a \in \Sigma_1$, and $(g_1 \times g_2)(a) = g_2(a)$, if $a \in \Sigma_2$. Obviously, $g_1 \times g_2$ is a square-free homomorphism, if g_1 and g_2 are square-free. Hence, $h_6^{(1)} \times \cdots \times h_6^{(r)}$ is a square-free homomorphism from $\{a_1, \ldots, a_{6r}\}^*$ into $\{a_1, \ldots, a_{3r}\}^*$. Now the repeated composition (of depth p) $h_6^{(1)} \circ \ldots \circ (h_6^{(1)} \times \cdots \times h_6^{(r)})$ is a square-free homomorphism from Σ^* into $\{a_1, a_2, a_3\}^*$. □

Remark. It should be noted that the homomorphisms from Theorem 4 are the shortest uniformly growing square-free homomorphisms from alphabets up to six letters into $\{a, b, c\}^*$. Furthermore, for $n \leq 3$, every uniformly growing square-free homomorphism from $\{a_1, \ldots, a_n\}^*$ into $\{a, b, c\}^*$ equals h_1, h_2, h_3 or h_3' up to a renaming of the letters and a reversal of the strings, where $h_3'(a_1) = abcacbcabac$, $h_3'(a_2) = abcbacabacb$, and $h_3'(a_3) = abcbacbcacb$.

The fact that these homomorphisms are the shortest square-free homomorphisms of their kind was checked by a PL/1 computer program, which was run on the IBM 370-165 of the RHRZ Bonn, and consumed 30 min CPU time.

The program first generated all square-free strings of length n over $\{a, b, c\}$. Then it exhaustively searched all k-tuples of these strings, which are compatible with each other according to the conditions of Theorem 2. These strings can be used as the homomorphic images of k letters.

The existence of square-free homomorphisms, which properly reduce arbitrary alphabets to three letter alphabets, can be used to show that the set of square-free strings over a three (and more) letter alphabet is exponentially dense.

Definition. For a language L and $n \geq 0$ let $\Pi_L(n)$ denote the number of strings of length n in L. Π_L is called the *density function* of L.

Theorem 5. *The set of square-free strings over a three letter alphabet is exponentially dense, i.e., there exist constants* $c_1, c_2 > 1$ *such that for every* $n \geq 2$

$$6 \cdot c_1^n \leq \Pi_{\mathrm{FREE}_{3(<2)}}(n) \leq 6 \cdot c_2^n.$$

Proof. There are 1172 square-free strings of length 24 over $\{a, b, c\}$ beginning with ab. Hence, each square-free string uv with $|v| = 2$ has at most 1172 square-free extensions $uv\alpha$ with $|\alpha| = 22$. Since there are 6 square-free strings of length two, $\Pi_{\mathrm{FREE}_{3(<2)}}(n) \leq 6 \cdot c_2^{n-2}$, where $c_2 = 1172^{1/22} \leq 1.38$. To establish the lower bound let w be a square-free string of length $l>0$ over $\{a, b, c\}$, which exists by Example 1. Define a finite substitution by $\tau(a)=\{a, a'\}$, $\tau(b)=\{b, b'\}$ and $\tau(c)=\{c, c'\}$. Then each string $x \in \tau\{w\}$ is a square-free string over $\{a, a', b, b', c, c'\}$ and $h_6(x) \in \{a, b, c\}^*$ is a square-free string of length $22l$. $\Pi_{\tau\{w\}}(l)=2^l$. For every $y_1, y_2 \in \tau\{w\}$ with $y_1 \neq y_2$, $h_6(y_1)$ and $h_6(y_2)$ have different prefixes of length m with $22(l-1) \leqslant m \leqslant 22l$. This follows from the fact that h_6 is injective, which is a consequence of the square-freeness. Hence, for every $n>2$ there exist at least $6 \cdot 2^l$ square-free strings of length n in $\{a, b, c\}^*$, where $l=\lceil n/22 \rceil$. Thus $\Pi_{\mathrm{FREE}_{3(<2)}}(n) \geqslant 6 \cdot 2^{n/22} \geqslant 6 \cdot 1.032^n$. □

For $n=1, 2, \ldots, 24$ that actual numbers of square-free strings over $\{a, b, c\}$ are as follows: 3, 6, 12, 18, 30, 42, 60, 78, 108, 144; 204, 264, 342, 456, 618, 798, 1044, 1392, 1830, 2388; 3180, 4146, 5418, 7032. This sequence suggests that the density function of the set of square-free strings over a three letter alphabet grows at least as 1.3^n, which means that our upper bound is better than our lower bound. Important is that both bounds are exponential.

Notice that the iteration of a single square-free homomorphism defines only sparse languages of density $O(n \cdot \log n)$ as it has been shown in [4].

For cube-free strings over two letter alphabets we proceed in a similar way, improving again a result by Bean et al. [1].

Theorem 6. *For every alphabet Σ there exists a uniformly growing cube-free homomorphism h from Σ^* into $\{a, b\}^*$.*

Proof. Let $\Sigma=\{a_1, a_2, \ldots, a_n\}$ and define homomorphisms $h_1, \ldots, h_4$ by the table

	h_1	h_2	h_3	h_4
a_1	ab	ab	$aababb$	$aabaabbab$
a_2		ba	$aabbab$	$aababaabb$
a_3			$abbaab$	$aabbababb$
a_4				$abbaababb$

Obviously, h_1 is cube-free, and h_2 is cube-free by Theorem 3. The homomorphism h_3 is cube-free on all cube-free strings of length four, but h_3 does not satisfy the 'substring property' from [1], which requires that $h(ab)=uh(c)v$ implies that $u=\lambda$ and $a=c$, or $v=\lambda$ and $b=c$ for all letters a, b, c. Here the homomorphic images of the letters are of the form xy, yx and xz, and yx is a substring of $xyxz$. However, each of the strings x, y, z of length three is unique, and the prefix $aabb$ of $h_3(a_2)$ serves as a separation marker so that its occurrence in a string $h_3(w)$ uniquely determines the occurrence of a_2 in w.

Suppose that $h_3(w)=puuuq$ has a cube, and that w is of minimal length. If $aabb$ occurs in u, then $w=\alpha u_1 a_2 u_2 \beta u_1 a_2 u_2 \beta u_1 a_2 u_2 \gamma$ with $\alpha, \beta, \gamma \in \{a_1, a_2, a_3\} \cup \{\lambda\}$, and a simple case analysis as in Theorem 2 shows $\alpha=\beta$ or $\gamma=\beta$, such that w has a cube. Conversely, if *aabb* does not occur in u, then a_2 occurs at most as the first or the last letter of w. By the uniqueness of x, y and z, $w=a_2va_3va_3va_3$, $w=a_1va_1va_1va_2$ or $w=vvv$ with $v\in\{a_1,a_3\}^*$. In each case, w has a cube. Hence, h_3 is cube-free.

The homomorphism h_4 is cube-free on all cube-free strings of length four. But h_4 does not satisfy the "substring property", since $h_4(a_3)$ occurs as a substring of $h_4(a_1a_4)$. However, *aabaa*, *ababa* and *aababb* serve as separation markers and uniquely identify $h_4(a_1)$, $h_4(a_2)$, and $h_4(a_4)$, respectively. Now an analysis as above for h_3 shows that h_4 is a cube-free homomorphism.

If the alphabet contains more than four letters, then repeated compositions of (extensions of) the homomorphism h_4 as in Theorem 4 define a cube-free homomorphism from Σ^* into $\{a,b\}^*$. □

Remark. It should be noted that the homomorphisms from Theorem 6 are the shortest uniformly growing cube-free homomorphisms from alphabets up to four letters into two letter alphabets. Furthermore, the homomorphisms h_1, h_2, and h_3, h_3' and h_3'' are unique up to a renaming of the letters or a reversal of the strings, where $h_3'(a_1)=aabbab$, $h_3'(a_2)=abbaab$, $h_3'(a_3)=babaab$, and $h_3''(a_1)=aabbab$, $h_3''(a_2)=baabab$, $h_3''(a_3)=babaab$.

Using the fact that there are 1251 cube-free strings of length 18 in $\{a,b\}^*$ which begin with a and the homomorphism h_4 from Theorem 6 we obtain that the set of cube-free strings over two letter alphabets is exponentially dense.

Theorem 7. *The set of cube-free strings over a two letter alphabet is exponentially dense, i.e., there exist constants d_1, d_2 such that for every $n>0$*

$$2\cdot d_1^n\leqslant \Pi_{\mathrm{FREE}_2(<3)}(n)\leqslant 2\cdot d_2^n.$$

Here, $d_1\geqslant 2^{1/9}\geqslant 1.08$ and $d_2\leqslant 1251^{1/17}\leqslant 1.522$.

4. Repetitive thresholds

From the aforesaid we know that there exist infinitely many square-free strings over a three letter alphabet and infinitely many cube-free strings over a two letter alphabet. However, if the size of the alphabets is reduced by one or if the repetitive power is decreased e.g., to $\frac{3}{2}$, then the sets of kth power-free strings are finite. Thus there is the problem of determining the least repetitive power k such that for every n letter alphabet Σ, $\mathrm{FREE}_\Sigma(<k)$ is finite, and $\mathrm{FREE}_\Sigma(\leqslant k)$ is infinite, and of establishing lower bounds on the density of $\mathrm{FREE}_\Sigma(\leqslant k)$. We cannot solve these problems here, since our techniques from above fail.

Definition. For every $n \geq 1$ the *repetitive threshold* $\mathrm{RT}(n)$ is the rational number k such that for every n letter alphabet Σ $\mathrm{FREE}_\Sigma(<k)$ is finite and $\mathrm{FREE}_\Sigma(\leq k)$ is infinite.

Since $\mathrm{FREE}_\Sigma(<\mathrm{RT}(n))$ is finite for every n letter alphabet Σ there is an upper bound on the length of the strings in $\mathrm{FREE}_\Sigma(<\mathrm{RT}(n))$. Hence the set of exactly $\mathrm{RT}(n)$th power-free strings $\mathrm{FREE}_\Sigma(=\mathrm{RT}(n))$ is infinite.

The notion of a repetitive threshold is due to Déjean [3]. Our definition is based on powers of strings, whereas Déjean has considered the relationship between the length of some strings u and v such that uvu is the kth power of uv for some k. For alphabets up to three letters the repetitive thresholds are known.

Theorem 8.

$$\mathrm{RT}(1)=\infty, \qquad \mathrm{RT}(2)=2, \qquad \mathrm{RT}(3)=\tfrac{7}{4},$$

and

$$1<\mathrm{RT}(n+1)\leq \mathrm{RT}(n) \quad \textit{for every positive integer } n.$$

Proof. $\mathrm{FREE}_1(\leq k)$ contains $[k+1]$ elements, which implies that $\mathrm{RT}(1)=\infty$. $\mathrm{FREE}_{\{a,b\}}(<2)=\{\lambda, a, b, ab, ba, aba, bab\}$, whereas $\mathrm{FREE}_{\{a,b\}}(\leq 2)$ is infinite, since it contains the set of prefixes of the infinite sequence generated by the 'Thue' homomorphism from Example 1. Finally, $\mathrm{FREE}_3(<\frac{7}{4})$ contains 3196 strings of length up to 38 as shown by Déjean [3], and her homomorphism from Example 2 is weakly $\frac{7}{4}$th power-free and defines an infinite weakly $\frac{7}{4}$th power-free string so that $\mathrm{FREE}_3(\leq\frac{7}{4})$ is infinite. Finally, it is obvious that the sequence of repetitive thresholds decreases, when the alphabets grow, and that $\mathrm{RT}(n)>1$. □

Remark. For $n \geq 4$ the repetitive thresholds are unknown. There is good evidence that $\mathrm{RT}(4)=\frac{7}{5}$ and $\mathrm{RT}(n)=n/(n-1)$ for $n \geq 5$, as proposed by Déjean [3]. $\mathrm{FREE}_4(<\frac{7}{5})$ is finite and consists of 236 345 strings of length up to 121, and every string of length $n+2$ over a n letter alphabet has a $n/(n-1)$th power, so that $\mathrm{FREE}_n(<n/(n-1))$ is finite. Hence, one condition of a repetitive threshold is satisfied by these values. However, we do not know whether they satisfy the second condition, too, since infinite or infinitely many weakly kth power-free strings have not yet been found in these cases. The following results show that new techniques are needed to determine the repetitive thresholds.

Theorem 9. *Let Σ and Δ be n letter alphabets and let h be a weakly* $\mathrm{RT}(n)$*th power-free homomorphism from Σ^* into Δ^*. Then $w \in \mathrm{FREE}_\Sigma(=\mathrm{RT}(n))$ implies $h(w) \in \mathrm{FREE}_\Delta(=\mathrm{RT}(n))$.*

Proof. From $\mathrm{RT}(2)=2$ we obtain $w \in \mathrm{FREE}_\Sigma(=2)$ if and only if $w=uxxv$. Then

$h(w) \in \text{FREE}_\Delta(=2)$. Consider $n \geq 3$. Since $\text{RT}(n) \leq \frac{7}{4}$ by Theorem 8,

$$\text{FREE}_\Sigma(=\text{RT}(n)) = \{uxyxv \in \Sigma^* \mid |y| = t \cdot |x|,\ t = (2-\text{RT}(n))/(\text{RT}(n)-1)\}.$$

Suppose that $h(w) \notin \text{FREE}_\Delta(=\text{RT}(n))$ for some $w \in \text{FREE}_\Sigma(=\text{RT}(n))$, i.e., $h(w) \in \text{FREE}_\Delta(<\text{RT}(n))$. Then $|h(y)| > t \cdot |h(x)|$, which implies that $|h(a)| \neq |h(b)|$, $\#_c(y) < t \cdot \#_c(x)$, and $\#_d(y) > t \cdot \#_d(y)$ for some letters $a, b, c, d \in \Sigma$, where $\#_c(y)$ denotes the number of occurrences of c in y. Consider $w' = u'x'y'x'v' \in \text{FREE}_\Sigma(=\text{RT}(n))$, which is obtained from $w = uxyxv$ by a renaming of the letters such that $|h(x')|$ is maximal among all strings so obtained. Then $t \cdot |h(x')| - |h(y')| > t \cdot |h(x)| - |h(y)|$, which implies $h(w') \notin \text{FREE}_\Delta(\leq \text{RT}(n))$, and h is not weakly $\text{RT}(n)$th power-free. □

From the proof of Theorem 9 we obtain that every weakly $\text{RT}(n)$th power-free homomorphism is either length uniform or the letters are uniformly distributed in each exactly $\text{RT}(n)$th power-free string $uxyxv$, such that $\#_a(y)/\#_a(x) = (2-\text{RT}(n))/(\text{RT}(n)-1)$ for all letters a and $xyx \in \text{FREE}_\Sigma(=\text{RT}(n))$.

Theorem 10. *If* $\text{RT}(4) = \frac{7}{5}$ *and* $\text{RT}(n) = n/(n-1)$ *for* $n \geq 5$, *then every weakly* $\text{RT}(n)$*th power-free homomorphism over n letter alphabets* Σ *and* Δ *is length uniform for every* $n \geq 3$.

Proof. Assume the contrary. Let $a, b \in \Sigma$ with $|h(a)| = \max\{|h(c)| \mid c \in \Sigma\}$, and $|h(b)| = \min\{|h(c)| \mid c \in \Sigma\}$. Thus $|h(a)| > |h(b)|$. Now $abcbabc \in \text{FREE}_\Sigma(=\frac{7}{4})$, but $h(abcbabc) \notin \text{FREE}_\Delta(=\frac{7}{4})$, $abcdbacbdcabcd \in \text{FREE}_\Sigma(=\frac{7}{5})$, but $h(abcdbacbdcabcd) \notin \text{FREE}_\Delta(=\frac{7}{5})$, and for $n \geq 5$, $a_1a_2 \ldots a_{n-1}a_1 \in \text{FREE}_\Sigma(=n/(n-1))$, but $h(a_1a_2 \ldots a_{n-1}a_1) \notin \text{FREE}_\Delta(=n/(n-1))$ with $a_1 = a$ and $a_2 = b$. □

Another restriction on weakly $\text{RT}(n)$th power-free homomorphisms is the following:

Theorem 11. *Let h be a weakly* $\text{RT}(n)$*th power-free homomorphism over n letter alphabets* Σ *and* Δ. *If* $h(a) = cu$ *and* $h(b) = dv$ $(h(a) = uc,\ h(b) = vd)$ *with* $a, b, c, d \in \Sigma$, *then* $a \neq b$ *implies* $c \neq d$. *Thus the beginning and the end of the homomorphic images of different letters must be different.*

Proof. Assume the contrary and let $h(a) = cu$ and $h(b) = cv$ with $a \neq b$. For $n = 2$, $h(aab) = cucucv$ is not weakly 2nd power-free, and thus contradicts the assumption. Consider $n \geq 3$ and $\text{RT}(n) \leq \frac{7}{4}$. Since $\text{FREE}_\Sigma(=\text{RT}(n))$ is infinite and invariant under a renaming of the letters there exists $xyxb \in \text{FREE}_\Sigma(=\text{RT}(n))$ with $|y| = t \cdot |x|$, where $t = (2-\text{RT}(n))/(\text{RT}(n)-1)$. Now $xyz \in \text{FREE}_\Sigma(=\text{RT}(n))$, $h(xyz) \in \text{FREE}_\Delta(=\text{RT}(n))$, and $y \neq bz$, i.e., $y = az$ by renaming. Then $h(xyxb) \notin \text{FREE}_\Delta(\leq \text{RT}(n))$, which contradicts the assumption that h is weakly $\text{RT}(n)$th power-free. □

As a consequence we obtain:

Theorem 12. *There exists no weakly* RT(n)*th power-free homomorphism from* Σ^* *into* Δ^*, *if* Σ *is an m letter alphabet,* Δ *is an n letter alphabet, and* $m>n$.

Theorem 12 shows that the technique employed in Theorem 5 and in Theorem 7 cannot be used to show that there exist exponentially many weakly RT(n)th power-free strings of each length. In fact, the density function of the set of weakly square-free strings over two letter alphabets is not strictly increasing (it makes some zig-zags; see $n=24-27$) and seems to grow slower than $2^{\sqrt{n}}$, as the first 95 values of the density function may suggest.

> 2, 4, 6, 10, 14, 20, 24, 30, 36, 44; 48, 60, 60, 72, 82, 88, 96, 112, 120; 120, 136, 148, 164, 152, 154, 148, 162, 176, 190; 196, 210, 216, 224, 228, 248, 272, 284, 296, 300; 296, 320, 332, 356, 356, 376, 400, 416, 380, 382; 376, 382, 356, 374, 392, 410, 432, 458, 464, 486; 476, 498, 500, 522, 528, 540, 548, 568, 560, 592; 592, 620, 660, 688, 688, 722, 724, 740, 724, 724; 716, 748, 788, 824, 816, 856, 868, 880, 868, 912; 908, 960, 976, 1008, 1000.

If the repetitive threshold is less than $\frac{3}{2}$, which is likely to be true for alphabets with at least four letters, the situation is even worse.

Theorem 13. *If* $\mathrm{RT}(n)<\frac{3}{2}$, *then there exists no growing weakly* RT(n)*th power-free homomorphism from* Σ^* *into* Δ^*, *where* Σ *and* Δ *are n letter alphabets.*

Proof. Let h be a homomorphism from Σ^* into Δ^* and assume $n\geqslant 4$, since $\mathrm{RT}(3)=\frac{7}{4}$. By Theorem 11, $h(a)=\alpha(a)\gamma(a)\beta(a)$ for each letter a, where α and β are renamings (permutations) of Δ. Assume that $h(a)=a_1a_2u$ with $a_1, a_2\in\Sigma$. From Theorem 11 we obtain $\beta(b)\neq a_1$ for all $b\in\Sigma$ with $b\neq a$ and $h(c)=va_2$ for some $c\in\Sigma$. Thus $h(ca)=va_2a_1a_2u$ and h is not weakly RT(n)th power-free. □

By Theorem 13 it is impossible to define an infinite weakly RT(n)th power-free string using weakly RT(n)th power-free homomorphisms, if $\mathrm{RT}(n)<\frac{3}{2}$. Since RT($n$) is assumed to be less than $\frac{3}{2}$ for $n\geqslant 4$, this tool fails to work for determining the values of the repetitive thresholds. Nevertheless, an infinite weakly RT(n)th power-free string may be definable by the iteration of a homomorphism, which is weakly RT(n)th power-free only on its sequence of strings of the form $h^i(w)$ for some axiom w and all $i\geqslant 1$, i.e., a weakly RT(n)th power-free DOL system may exist (see [4]). In the case of four letter alphabets and $\mathrm{RT}(n)<\frac{3}{2}$ we disprove this assumption for uniformly growing homomorphisms. So there is no hope to determine the value of RT(4) easily.

Lemma. *Let $\Sigma = \{a, b, c, d\}$ and $w \in \mathrm{FREE}_\Sigma(=k)$ with $k < \frac{3}{2}$.*

(i) *If the letter a does not occur in w, then $|w| \leq 4$.*

(ii) *If the string ab does not occur in w as a substring, then $|w| \leq 22$.*

Proof. The longest string satisfying (i) is, e.g. *bcdb*, which proves (i). All strings in $\mathrm{FREE}_\Sigma(=k)$, which do not contain *ab* as a substring and begin with *ac* are shown in Fig. 1. By the symmetry of *c* and *d*, a similar set of strings is obtained for the prefix *ad*. All strings so obtained are no longer than 22, so that (i) implies $|w| \leq 26$ for all strings satisfying (ii). In fact if *w* satisfies (ii) and begins with *b*, *c*, or *d*, then $|w| \leq 22$. $\square$

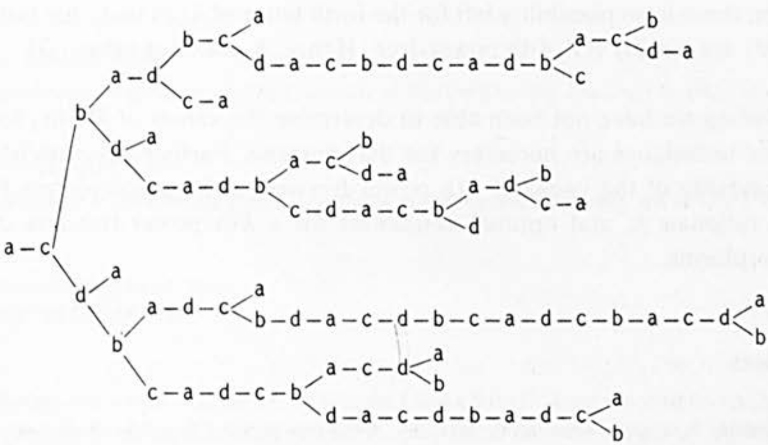

Fig. 1.

Notice that the particular strings *a* and *ab* are avoided in the strings of the above lemma, whereas substrings such as *b* or *ba* or *bc* may occur. Thus the notion of avoidable used here is more restrictive than the one in [1]. It coincides with Thue's notion of avoiding *aca* and *bcb* in square-free strings over $\{a, b, c\}$. See [9, Satz 4 and Satz 10].

Theorem 14. *If* $\mathrm{RT}(4) < \frac{3}{2}$, *then there exists no uniformly growing weakly* $\mathrm{RT}(4)$*th power-free homomorphism h on a four letter alphabet such that $h^i(w)$ is weakly* $\mathrm{RT}(4)$*th power-free for all $i \geq 1$ and some axiom w.*

Proof. Let $\Sigma = \{a, b, c, d\}$ and consider $i \geq 1$ such that $|h^i(a)| \geq 23$. By the previous Lemma, $h^i(a)$ contains all substrings pq with $p, q \in \Sigma$. Hence $h(pq) \in \mathrm{FREE}_\Sigma(\leq \mathrm{RT}(4))$ for all letters p, q with $p \neq q$. Since $\mathrm{FREE}_\Sigma(<\mathrm{RT}(4))$ is finite there is an upper bound K on the length of its strings. Thus for all $i > K/|h(a)|$, $h^i(a)$ is in $\mathrm{FREE}_\Sigma(=\mathrm{RT}(4))$, and if $|h^i(a)| > K+2$, then $h^i(a) = uxyxv$ with $u, v \neq \lambda$, and xyx is the exact $\mathrm{RT}(4)$th power of xy. Since h is growing appropriate i's exist.

If h is weakly RT(4)th power-free on its images $h^i(w)$, then u and y end with different letters and the last letters of the homomorphic images of these letters are different. Otherwise, $uxyxv$ resp. $h(uxyxv)$ is not weakly RT(4)th power-free. By symmetry v and y begin with different letters and the homomorphic images of these letters begin with different letters. We now try to fix the homomorphism h, and for our convenience we assume $y = ay'$, $v = bv'$, $h(a) = a\alpha$ and $h(b) = b\beta$. Since $h(ca)$, $h(cb)$, $h(da)$ and $h(db)$ are weakly RT(4)th power-free we obtain that $h(c) = \gamma c$ and $h(d) = \delta d$ or $h(c) = \gamma d$ and $h(d) = \delta c$. Simply assume the first case. Then we can determine the second letters of the homomorphic images and obtain $h(a) = ab\alpha'$, $h(b) = ba\beta'$, $h(c) = \gamma' dc$ and $h(d) = \delta' cd$, where α', β', γ', $\delta' \neq \lambda$. Continuing in this way $h(a)$ may begin with abc or abd, but $h(ab)$ forces $|h(a)| \geq 4$. However, there is no possibility left for the forth letter of $h(a)$ such that both $h(ca)$ and $h(da)$ are weakly RT(4)th power-free. Hence, h does not exist. □

Concluding we have not been able to determine the values of $RT(n)$ for $n \geq 4$, since new techniques are necessary for that purpose. Further open problems are the decidability of the (weakly) kth power-freeness of homomorphisms for non-integral rationals k, and optimal conditions for a kth power-freeness check of homomorphisms.

References

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

[2] J. Berstel, Sur les mots sans carré définis par un morphisme, *Proc. 6th International Colloquium Automata, Languages, and Programming*, Lecture Notes in Computer Science **71** (Springer, Berlin, 1979) 16–29.

[3] F. Déjean, Sur un théorème de Thue, *J. Combinatorial Theory* **13** (1972) 90–99.

[4] A. Ehrenfeucht and G. Rozenberg, On the subword complexity of square-free DOL languages, *Theoret. Comput. Sci.* **16** (1981) 339–342.

[5] M.A. Harrison, *Introduction to Formal Language Theory* (Addison-Wesley, Reading, MA, 1978).

[6] R. Lyndon and M.P. Schützenberger, The equations $a^M = b^N c^P$ in a free group, *Michigan Math. J.* **9** (1962) 289–298.

[7] A. Ostrowski, *Vorlesung über Differential- und Integralrechnung*, Erster Band (Birkhäuser Verlag, Stuttgart, 1960).

[8] A. Salomaa, Morphisms on free monoids and language theory, in: R. V. Book, Ed., *Formal Language Theory: Perspectives and Open Problems* (Academic Press, New York, 1980) 141–166.

[9] A. Thue, Über unendliche Zeichernreihen, *Norske Vid. Selsk. Skr. I, Mat.-Nat. Kl. Christiana* **7** (1906) 1–22.

[10] A. Thue, Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen, *Norske Vid. Selsk. Skr. I, Mat.-Nat. Kl. Christiana* **1** (1912) 1–67.