A Note on the Maximum Number of k-Powers in a Finite Word

Shuo Li[a]   Jakub Pachocki[b]   Jakub Radoszewski[c]

Abstract

A *power* is a concatenation of k copies of a word u, for a positive integer k; the power is also called a *k-power* and k is its *exponent*. We prove that for any $k \geqslant 2$, the maximum number of different non-empty k-power factors in a word of length n is between $\frac{n}{k-1} - \Theta(\sqrt{n})$ and $\frac{n-1}{k-1}$. We also show that the maximum number of different non-empty power factors of exponent at least 2 in a length-n word is at most $n - 1$. Both upper bounds generalize the recent upper bound of $n - 1$ on the maximum number of different square factors in a length-n word by Brlek and Li (2022).

Mathematics Subject Classifications: 68R15

1   Introduction

Let k be an integer greater than 1. The *k-power* (or simply the *power*) of a word u is a word of the form $\underbrace{uu \dots u}_{k \text{ times}}$. Here k is called the *exponent* of the power. We consider only powers of non-empty words. A factor (subword) of a word is its fragment consisting of a number of consecutive letters. In this paper, we investigate the bounds for the maximum number of different k-power factors in a word of length n. This subject is one of the fundamental topics in combinatorics on words [9]. For any pair of positive integers (n, k) with $k > 1$, let $N(n, k)$ denote the maximum number of different non-empty k-powers that can appear as factors of a word of length n. For 2-powers (squares), the bounds for $N(n, 2)$ were studied by many authors; see [3, 4, 6, 2, 10, 1]. The best known lower bound from [3] and a very recent upper bound from [1] match up to sublinear terms:

$$n - o(n) \leqslant N(n, 2) \leqslant n - 1.$$

[a]Laboratoire de Combinatoire et d'Informatique Mathématique, Université du Québec à Montréal (`shuo.li.ismin@gmail.com`).

[b]Open AI, San Francisco, CA, U.S.A. (`jakub@openai.com`).

[c]Institute of Informatics, University of Warsaw, Poland (`jrad@mimuw.edu.pl`).

Actually, one can check that the lower bound from [3] is of the form $n - \Theta(\sqrt{n})$. For $k = 3$, it was proved in [5] that

$$\frac{1}{2}n - 2\sqrt{n} \leqslant N(n,3) \leqslant \frac{4}{5}n.$$

More generally, for $k \geqslant 3$, it was studied in [7] and proved that

$$N(n,k) \leqslant \frac{n-1}{k-2},$$

with the same notation as above. Further in [5] it was shown that the maximum number of different factors of a word of length n being powers of exponent *at least* 3 is $n - 2$.

In this article, we generalize the methods provided in [1] and [5] to give an upper and a lower bound for the number of different k-powers in a finite word. The main result is announced as follows:

Theorem 1. *Let k be an integer greater than 1. For any integer $n > 1$, let $N(n,k)$ denote the maximum number of different k-powers being factors of a word of length n. Then we have*

$$\frac{n}{k-1} - \Theta(\sqrt{n}) \leqslant N(n,k) \leqslant \frac{n-1}{k-1}.$$

We also show the following result. It implies, in particular, that a word that contains powers of exponent greater than 2 has fewer squares than $n - 1$.

Theorem 2. *The maximum number of different factors in a word of length n being powers of exponent at least 2 is $n - 1$.*

2 Preliminaries

Let us first recall the basic terminology related to words. By a *word* we mean a finite concatenation of symbols $w = w_1 w_2 \cdots w_n$, with n being a non-negative integer. The *length* of w, denoted $|w|$, is n and we say that the symbol w_i is at the *position i*. The set $\mathrm{Alph}(w) = \{w_i \mid 1 \leqslant i \leqslant n\}$ is called the *alphabet* of w and its elements are called *letters*. Let $|\mathrm{Alph}(w)|$ denote the cardinality of $\mathrm{Alph}(w)$. A word of length 0 is called the *empty word* and it is denoted by ε. *Concatenation* of two words u, v is denoted as uv.

A word u is called a *factor* of a word w if $w = pus$ for some words p, s; u is called a *prefix (suffix)* of w if $p = \varepsilon$ ($s = \varepsilon$, respectively). The set of all factors of a word w is denoted by $\mathrm{Fac}(w)$.

Two words u and v are *conjugate* when there exist words x, y such that $u = xy$ and $v = yx$. The conjugacy class of a word v is denoted by $[v]$. If $v = v_1 v_2 \cdots v_m$ is a word, then for any $i \in \{1, \ldots, m\}$, we define $v_p(i) = v_1 v_2 \cdots v_i$ and $v_s(i) = v_{i+1} v_{i+2} \cdots v_m$. Thus, $[v] = \{v_s(i) v_p(i), i = 1, 2, \ldots, m\}$.

For any positive integer k, we define the *k-power* (or simply a *power*) of a word u to be the concatenation of k copies of u, denoted by u^k. Here k is the *exponent* of the power. In

particular, $\varepsilon^k = \varepsilon$ for any natural number k. A non-empty word w is said to be *primitive* if it is not a power of another word, that is, if $w = u^k$ implies $k = 1$. For any non-empty word w, there is exactly one primitive word u such that $w = u^k$ for integer $k \geqslant 1$; the word u is called the *primitive root* of word w, see [8]. Furthermore, two words that are conjugate are either both primitive or none of them is [8, Proposition 1.3.3]. For a given word w, let $N_k(w)$ denote the number of different non-empty k-power factors of w and $\mathrm{Prim}(w)$ denote all primitive factors of w.

For any word u and any rational number α, the α-*power* of u is defined to be $u^a u_0$ where u_0 is a prefix of u, a is the integer part of α, and $|u^a u_0| = \alpha|u|$. The α-power of u is denoted by u^α. If α is a rational number greater than 1 and there exists a word u such that $w = u^\alpha$, then the word w is said to have a *period* $|u|$.

3   Rauzy graphs of a finite word

In this section, we recall the notion of Rauzy graph and some results obtained in [1]. Let w be a word of length n. For any integer $\ell \in \{1, \ldots, n\}$, let $L_\ell(w)$ be the set of all length-ℓ factors of w. For any integer $\ell \in \{1, \ldots, n\}$, let the Rauzy graph $\Gamma_\ell(w)$ be an oriented graph whose set of vertices is $L_\ell(w)$ and the set of edges is $L_{\ell+1}(w)$ (here $L_{n+1}(w) = \emptyset$); an edge $e \in L_{\ell+1}(w)$ starts at the vertex u and ends at the vertex v, if u is a prefix and v is a suffix of e. Let us define $\Gamma(w) = \cup_{\ell=1}^n \Gamma_\ell(w)$.

Let $\Gamma_\ell(w)$ be a Rauzy graph of w. A sub-graph in $\Gamma_\ell(w)$ is called an *elementary circuit* if there are j distinct vertices $v_1, v_2, \ldots, v_j$ and j distinct edges $e_1, e_2, \ldots, e_j$ for some integer j, such that for each t with $1 \leqslant t \leqslant j - 1$, the edge e_t starts at v_t and ends at v_{t+1}, and for the edge e_j, it starts at v_j and ends at v_1; further, j is called the *size* of the circuit. The *small circuits* in the graph $\Gamma_\ell(w)$ are those elementary circuits whose sizes are no larger than ℓ.

Lemma and Notation 3 (Brlek and Li [1])**.** *Let w be a word and let $\Gamma_\ell(w)$ be a Rauzy graph of w for some $\ell \in \{1, \ldots, |w|\}$. Then for any small circuit C on $\Gamma_\ell(w)$, there exists a unique primitive word q, up to conjugacy, such that $|q| \leqslant \ell$ and the vertex set of C is $\left\{ p^{\frac{\ell}{|p|}} | p \in [q] \right\}$ and its edge set is $\left\{ p^{\frac{\ell+1}{|p|}} | p \in [q] \right\}$.*

Further, each small circuit can be identified by an associated primitive word q and an integer ℓ such that $\Gamma_\ell(w)$ is the Rauzy graph in which the circuit is located. Let each small circuit be denoted by $C(q, \ell)$ with the parameters defined as above.

Lemma 4 (Brlek and Li [1])**.** *Let w be a word. Then there are at most $|w| - |\mathrm{Alph}(w)|$ small circuits in $\Gamma(w)$.*

4   Upper bound for $N(n, k)$

Let w be a word and let $v \in \mathrm{Prim}(w)$. A factor $u \in \mathrm{Fac}(w)$ is said to be *in the class of factor* $v \in \mathrm{Fac}(w)$ if there is a (primitive) word $y \in [v]$ and an integer $p \geqslant 2$ such that

$u = y^p$. Let $\text{Class}_w(v)$ denote the set of all factors of w in the class of v. By $|\text{Class}_w(v)|$ we denote the cardinality of $\text{Class}_w(v)$.

For a factor v of w, let us define $m_w(v) = \max\{n | v^n \in \text{Fac}(w), n \in \mathbb{N}^+\}$. Now given $\text{Class}_w(v)$, let us define its *index* to be an integer $\text{Index}_w(v)$ such that $\text{Index}_w(v) = \max\{m_w(u) | u \in [v]\}$. From the definition, the elements in $\text{Class}_w(v)$ are all of the form $v_s(i) v^{j-1} v_p(i)$ with $1 \leqslant i \leqslant |v|$ and $1 \leqslant j \leqslant \text{Index}_w(v)$. By $\text{Prim}'(w)$ we denote the set of primitive words v such that v^n is in the class of v, where n is the index of this class. In other words,

$$\text{Prim}'(w) = \{v \in \text{Prim}(w) | v^{\text{Index}_w(v)} \in \text{Class}_w(v)\}.$$

For $v \in \text{Prim}'(w)$, let $\text{MaxPow}_w(v) = \{u^{\text{Index}_w(v)} | u \in [v]\} \cap \text{Fac}(w)$ and $\text{mp}_w(v)$ denote the cardinality of $\text{MaxPow}_w(v)$.

Example 5. Let $w = (00001)^3 2(00100)^3 2(01000)^3$ and $v = 00001$. Then $\text{Class}_w(v)$ has 8 elements:

$$\text{Class}_w(v) = \{(00001)^2, (00010)^2, (00100)^2, (01000)^2, (10000)^2,$$
$$(00001)^3, (00100)^3, (01000)^3\}.$$

In this case, we have $\text{Index}_w(v) = 3$, $\text{MaxPow}_w(v) = \{(00001)^3, (00100)^3, (01000)^3\}$, and $\text{mp}_w(v) = 3$ is the cardinality of $\text{MaxPow}_w(v)$. Moreover, the only words that are conjugate with v in $\text{Prim}'(w)$ are $00001, 00100, 01000$.

Lemma 6. *Let u and v be primitive words. If words u and v are conjugate, then $\text{Class}_w(u) = \text{Class}_w(v)$. Otherwise, classes $\text{Class}_w(u)$ and $\text{Class}_w(v)$ are disjoint.*

Proof. The first part of the statement is obvious. Assume to the contrary that $y \in \text{Class}_w(u) \cap \text{Class}_w(v)$ for primitive words u and v. This means that there exist words $u' \in [u]$ and $v' \in [v]$ and integers $k, t > 1$ such that $y = (u')^k = (v')^t$. However, in this case, u' and v' are powers of the same word (see [8, Proposition 1.3.1]). Moreover, since u' and v' are conjugate to primitive words, they are both primitive. Thus, $u' = v'$ and $\text{Class}_w(u) = \text{Class}_w(v)$. $\qquad\square$

In this section we give an upper bound for $N_k(w)$. The strategy is as follows: first, we compute the exact number of powers in each class $\text{Class}_w(v)$ of w; second, we prove that there exists an injection from $\cup_{v \in \text{Prim}'(w)} \text{Class}_w(v)$ to the set of small circuits in $\Gamma(w)$; third, we conclude by using the properties of Rauzy graphs introduced in the previous section.

Lemma 7. *Let w be a word and $v \in \text{Prim}'(w)$. If $\text{Index}_w(v) \geqslant 2$, then we have $|\text{Class}_w(v)| = |v|(\text{Index}_w(v) - 2) + \text{mp}_w(v)$. Further, we have*

$$\text{Class}_w(v) = \{u^k | u \in [v], 2 \leqslant k \leqslant \text{Index}_w(v) - 1\} \cup \text{MaxPow}_w(v).$$

Proof. We only need to prove that for any $u \in [v]$ and any integer k satisfying $2 \leqslant k \leqslant \text{Index}_w(v) - 1$, $u^k \in \text{Fac}(w)$. We can easily check that $u^k \in \text{Fac}(v^{k+1})$. However, from the hypothesis, $k+1 \leqslant \text{Index}_w(v)$ and $v^{\text{Index}_w(v)} \in \text{Fac}(w)$, thus, $v^{k+1} \in \text{Fac}(w)$. Consequently, $u^k \in \text{Fac}(w)$. $\qquad\square$

Example 9 below shows the main ideas from the proof of the following lemma for a concrete class.

Lemma 8. *Let w be a word and $v \in \mathrm{Prim}'(w)$ such that $|v| = \ell$ and $\mathrm{Index}_w(v) \geqslant 2$. For any integer t satisfying $1 \leqslant t \leqslant |\mathrm{Class}_w(v)|$, there exists a small circuit $C(v, t + \ell - 1)$ in the Rauzy graph $\Gamma_{t+\ell-1}(w)$. Hence, there exists a bijective function f_v which associates each word in $\mathrm{Class}_w(v)$ to a small circuit in the set $\{C(v, t + \ell - 1) | 1 \leqslant t \leqslant |\mathrm{Class}_w(v)|\}$.*

Proof. To prove the existence of the circuit $C(v, t+\ell-1)$ for any $t \in \{1, \ldots, |\mathrm{Class}_w(v)|\}$, it is enough to prove that

$$S_t(v) := \left\{ u^{\frac{t+\ell}{\ell}} \mid u \in [v] \right\} \subset \mathrm{Fac}(w).$$

If this is the case, then there exists a circuit in $\Gamma_{t+\ell-1}(w)$ such that its edge set is S_t; further, it can be identified by $C(v, t + \ell - 1)$.

If integers t', t satisfy $1 \leqslant t' < t \leqslant |\mathrm{Class}_w(v)|$, then each word in $S_{t'}(v)$ is a prefix of a word in $S_t(v)$. Hence, it is enough to prove that $S_t(v) \subset \mathrm{Fac}(w)$ for $t = |\mathrm{Class}_w(v)|$. From Lemma 7, we have $t = \ell(\mathrm{Index}_w(v) - 2) + \mathrm{mp}_w(v)$, so $\frac{t+\ell}{\ell} = \mathrm{Index}_w(v) - 1 + \frac{\mathrm{mp}_w(v)}{\ell}$.

Let $i = \mathrm{Index}_w(v)$ and $j = \mathrm{mp}_w(v)$. For any $u^{i-1+\frac{j}{\ell}} \in S_t(v)$, the word $u^{i-1+\frac{j}{\ell}} = u^{i-1}u_p(j)$ is a factor of the word $u_s(m)u^{i-1}u_p(m) = (u_s(m)u_p(m))^i$ for all $m \in \{j, \ldots, \ell\}$. Hence, there are at most $j-1$ distinct words y that are conjugate with u such that $u^{i-1+\frac{j}{\ell}} \notin \mathrm{Fac}(y^i)$. In particular, there are at most $j - 1$ distinct words $y^i \in \mathrm{MaxPow}_w(v)$ which do not contain $u^{i-1+\frac{j}{\ell}}$ as a factor. However, there are exactly j elements in $\mathrm{MaxPow}_w(v)$, so there exists at least one word in $\mathrm{MaxPow}_w(v)$ containing $u^{i-1+\frac{j}{\ell}}$. Thus, $S_t(v) \subset \mathrm{Fac}(w)$.

The existence of the bijective function f_v is from the fact that the cardinalities of $\mathrm{Class}_w(v)$ and $\{C(v, t + |v| - 1) | 1 \leqslant t \leqslant |\mathrm{Class}_w(v)|\}$ are the same. $\qquad\square$

Example 9. Let us consider the class from Example 5. All words in the set $S_8(v) = \left\{ u^{\frac{13}{5}} \mid u \in [v] \right\}$, for $v = \texttt{00001}$, are factors of w. For example, let us consider $u^{\frac{13}{5}} = (\texttt{00010})^{\frac{13}{5}} = \texttt{0001000010000} \in S_8(v)$. It is a factor of three words y^3, where u and y are conjugate:

$$(\texttt{10000})^3 = \texttt{10}\underline{\texttt{0001000010000}}$$
$$(\texttt{00001})^3 = \texttt{0}\underline{\texttt{0001000010000}}\texttt{1}$$
$$(\texttt{00010})^3 = \underline{\texttt{0001000010000}}\texttt{10}$$

and is not a factor of the two remaining such words. The set $\mathrm{MaxPow}_w(v)$ contains three words and indeed $u^{\frac{13}{5}}$ is a factor of y^3 for y being one of them, $y = \texttt{00001}$.

Now, since $S_8(v) \subset \mathrm{Fac}(w)$, we have $S_t(v) = \left\{ u^{\frac{5+t}{5}} \mid u \in [v] \right\} \subset \mathrm{Fac}(w)$, for all integer $1 \leqslant t \leqslant 8$. Moreover, each $S_t(v)$ is the edge set of the circuit $C(v, t + 4)$. Thus, there exists a bijective function from the class $\mathrm{Class}_w(v)$ to the set of circuits $\{C(v, t + \ell - 1) | 1 \leqslant t \leqslant |\mathrm{Class}_w(v)|\}$ since the cardinalities of these sets are both 8.

For a word w, let $\mathrm{Powers}(w)$ denote the set of all powers of exponent at least 2 that are factors of w, i.e. $\mathrm{Powers}(w) = \{u^t | t \geqslant 2, u^t \in \mathrm{Fac}(w)\}$.

Lemma 10. *There exists an injective function f from the set $\mathrm{Powers}(w)$ to the set of small circuits in $\Gamma(w)$.*

Proof. Each power factor of w of exponent at least 2 belongs to some class. Hence, $\mathrm{Powers}(w) = \cup_{v \in \mathrm{Prim}'(w)} \mathrm{Class}_w(v)$. For any $v \in \mathrm{Prim}'(w)$, from Lemma 8, there is a bijection f_v from $\mathrm{Class}_w(v)$ to $\{C(v, t + |v| - 1) | 1 \leqslant t \leqslant |\mathrm{Class}_w(v)|\}$. Let us define the function f as follows: for any $\mathrm{Class}_w(v)$, we set $f|_{\mathrm{Class}_w(v)} = f_v$ with f_v defined as above. This function is well defined by Lemma 6. Now we prove that f is injective. Let y, z be two powers such that $f(y) = f(z) = C(v, t + |v| - 1)$ for some v and t. In this case, y, z are both in $\mathrm{Class}_w(v)$. However, for a given class $\mathrm{Class}_w(v)$, f_v is bijective, thus $y = z$. $\square$

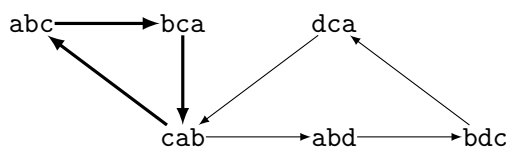The function f from Lemma 10 does not need to be a bijection; see Fig. 1.

Figure 1: Rauzy graph $\Gamma_3(w)$ of word $w = $ `abcabdcabc` (edge labels omitted) contains a small circuit (in bold) even though the word w is square-free.

Example 11. Let us consider a word $w = $ `10101001001000` (it is a prefix of the family of words considered in Section 5). Fig. 2 shows the Rauzy graphs $\Gamma_1(w), \ldots, \Gamma_8(w)$. The graph $\Gamma_8(w)$ is acyclic, so the remaining graphs $\Gamma_i(w)$, for $i \in \{9, \ldots, 14\}$, are acyclic as well. Small circuits in the Rauzy graphs are drawn in thick using different colors, depending on the class that they correspond to (as in the proof of Lemma 10):

- two blue circuits $C(0, i)$, $i \in \{1, 2\}$, corresponding to $\mathrm{Class}_w(0) = \{0^2, 0^3\}$,

- three green circuits $C(10, i)$, $i \in \{2, 3, 4\}$, corresponding to $\mathrm{Class}_w(10) = \{(10)^2, (01)^2, (10)^3\}$, and

- five red circuits $C(100, i)$, $i \in \{3, 4, 5, 6, 7\}$, corresponding to $\mathrm{Class}_w(100) = \{(100)^2, (010)^2, (001)^2, (100)^3, (010)^3\}$.

In total $|\mathrm{Powers}(w)| = 10$ and there are 10 small circuits in the Rauzy graphs. We note that not all circuits in this example are small; see the graphs $\Gamma_1(w)$ and $\Gamma_2(w)$.

We are ready to prove the upper bounds.

Lemma 12. *For any non-empty word w, $|\mathrm{Powers}(w)| \leqslant |w| - |\mathrm{Alph}(w)|$.*

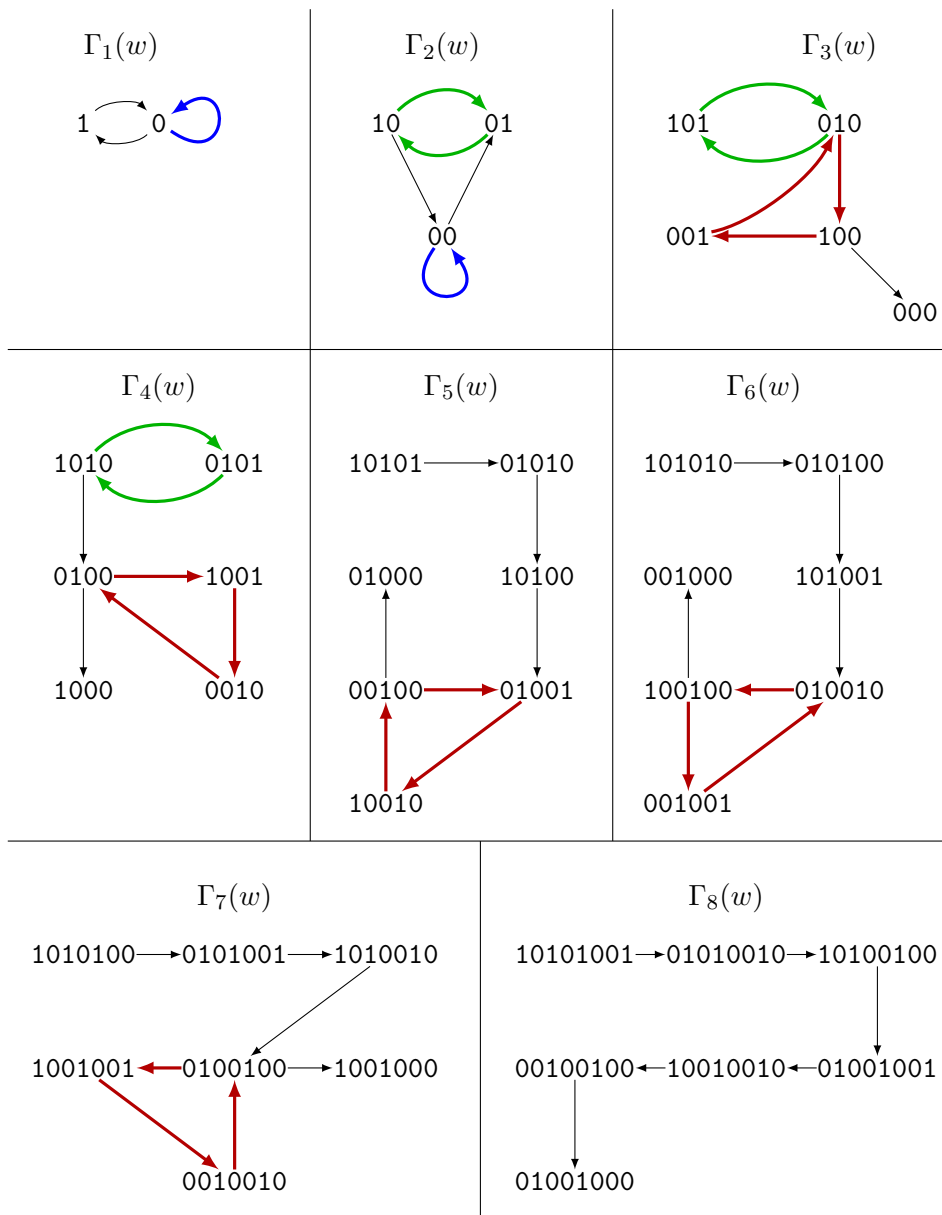*Proof.* It is a direct consequence of Lemmas 10 and 4. $\square$

Figure 2: Rauzy graphs $\Gamma_1(w), \ldots, \Gamma_8(w)$ of word $w = \mathtt{10101001001000}$ (edge labels omitted).

Theorem 2 follows directly from Lemma 12.

Theorem 13 (Upper bound).
Let k be an integer greater than 1. For any word w, we have

$$N_k(w) \leqslant \frac{|w| - |\mathrm{Alph}(w)|}{k-1}.$$

Consequently, for any integer $n \geqslant 1$, we have

$$N(n,k) \leqslant \frac{n-1}{k-1}.$$

Proof. To each k-power factor in $\mathrm{Powers}(w)$ we can assign at least $k-2$ powers in the set $\mathrm{Powers}(w)$ that are not k-powers. More precisely, if the k-power factor is v^{kp}, for positive integer p and primitive word v, then the words $v^{kp-1}, \ldots, v^{kp-k+2}$ are elements of $\mathrm{Powers}(w)$ and are not k-powers by uniqueness of primitive roots. (If $p > 1$, we could also assign v^{kp-k+1} to v^{kp}; however, for $p = 1$ this would be a 1-power.) Moreover, this way the sets of powers assigned to different k-powers are disjoint. By Lemma 12,

$$N_k(w) \leqslant \frac{|\mathrm{Powers}(w)|}{k-1} \leqslant \frac{|w| - |\mathrm{Alph}(w)|}{k-1}. \qquad \square$$

Example 14. For the word w from Example 11, we have $N_3(w) \leqslant |\mathrm{Powers}(w)|/2 = 5$. Actually, $N_3(w) = 4$ and $N_2(w) = 6$.

5 Lower bound for $N(n, k)$

We show a family of binary words which yields a lower bound of $\frac{n}{k-1} - \Theta(\sqrt{n})$ for the number of different factors which are k-powers, for an integer $k \geqslant 2$.

For integers $i \geqslant 1$ and $k \geqslant 2$ we denote

$$q_i^{(k)} = (\mathtt{10}^i)^{k-1}.$$

Let $r_m^{(k)}$ be the concatenation

$$r_m^{(k)} = q_1^{(k)} q_2^{(k)} \cdots q_m^{(k)} \mathtt{10}^m.$$

E.g., for $k = 2$, we obtain the family of words:

$$\mathtt{1010}, \ \mathtt{10100100}, \ \mathtt{1010010001000}, \ \mathtt{1010010001000010000}, \ldots$$

and for $k = 3$, the family:

$$\mathtt{101010}, \ \mathtt{1010100100100}, \ \mathtt{1010100100100010001000}, \ldots$$

Lemma 15. *The length of $r_m^{(k)}$ is $(k-1)\left(\frac{m^2}{2} + \frac{3m}{2}\right) + m + 1$.*

Proof. The length of $q_i^{(k)}$ is $(k-1)(i+1)$, so

$$|r_m^{(k)}| = \left(\sum_{i=1}^{m}(k-1)(i+1)\right) + m + 1 = (k-1)\left(\frac{m^2}{2} + \frac{3m}{2}\right) + m + 1. \qquad \square$$

Lemma 16. $N_k(r_m^{(k)}) \geqslant \frac{m^2}{2} + \frac{m}{2} + \left\lfloor\frac{m}{k}\right\rfloor$.

Proof. Let us note that for a positive integer i, the concatenation $0^{i-1}q_i^{(k)}10^i = 0^{i-1}(10^i)^k$ contains as factors all the k-powers that are conjugate with the k-power $(0^i1)^k$ that are different from this k-power. Let us note that this concatenation is a factor of $r_m^{(k)}$ for each $i \in \{1,\ldots,m\}$. Indeed, for $i \in \{1,\ldots,m\}$, the factor $q_i^{(k)}$ in $r_m^{(k)}$ is preceded by 0^{i-1} (for $i = 1$ this is the empty string, and otherwise it is a suffix of $q_{i-1}^{(k)}$) and followed by 10^i (for $i < m$ it is a prefix of $q_{i+1}^{(k)}$, and for $i = m$ it is a suffix of $r_m^{(k)}$).

Additionally, in $r_m^{(k)}$ there are $\left\lfloor\frac{m}{k}\right\rfloor$ unary k-powers $0^k, 0^{2k}, \ldots$ In total we obtain

$$\left(\sum_{i=1}^{m} i\right) + \left\lfloor\frac{m}{k}\right\rfloor = \frac{m^2}{2} + \frac{m}{2} + \left\lfloor\frac{m}{k}\right\rfloor$$

k-powers, all pairwise different. $\qquad \square$

Theorem 17 (Lower bound).
Let $k \geqslant 2$ be an integer. For infinitely many positive integers n there exists a word w of length n for which $N_k(w) > \frac{n}{k-1} - 2.2\sqrt{n}$.

Proof. Due to Lemmas 15 and 16, for any word $r_m^{(k)}$ we have:

$$\frac{|r_m^{(k)}|}{k-1} - N_k(r_m^{(k)}) \leqslant \frac{m^2}{2} + \frac{3m}{2} + \frac{m+1}{k-1} - \frac{m^2}{2} - \frac{m}{2} - \left\lfloor\frac{m}{k}\right\rfloor$$

$$= m + \frac{m+1}{k-1} - \left\lfloor\frac{m}{k}\right\rfloor \leqslant m + \frac{m+1}{k-1} - \frac{m-k+1}{k}$$

$$= m + \frac{m+1}{k(k-1)} + 1 \leqslant m + \frac{m+1}{2} + 1 = \frac{3}{2}m + \frac{3}{2}.$$

This value is smaller than $c\sqrt{|r_m^{(k)}|}$ for $c^2 \geqslant \frac{9}{2}$; indeed, in this case, we have:

$$\left(\frac{3}{2}m + \frac{3}{2}\right)^2 = \frac{9}{4}m^2 + \frac{9}{2}m + \frac{9}{4} < c^2\left(\frac{1}{2}m^2 + \frac{5}{2}m + 1\right) \leqslant c^2|r_m^{(k)}|.$$

Hence, for $c \geqslant 2.2$ we conclude that:

$$\frac{|r_m^{(k)}|}{k-1} - N_k(r_m^{(k)}) < c\sqrt{|r_m^{(k)}|} \quad \Rightarrow \quad N_k(r_m^{(k)}) > \frac{|r_m^{(k)}|}{k-1} - c\sqrt{|r_m^{(k)}|}. \qquad \square$$

Note. For $k = 2$ we obtain a family of words containing $n - o(n)$ different squares that is simpler than the example by Fraenkel and Simpson [3]: we concatenate the words $q_i^{(2)} = \mathtt{10}^i$ whereas they concatenate the words $q_i' = \mathtt{0}^{i+1}\mathtt{10}^i\mathtt{10}^{i+1}\mathtt{1}$.

Proof of Theorem 1. It is a direct consequence of Theorems 13 and 17. □

Acknowledgements

References

[1] S. Brlek and S. Li: On the number of squares in a finite word (2022), arXiv:2204.10204.

[2] A. Deza, F. Franek, and A. Thierry: How many double squares can a string contain? *Discret. Appl. Math.*, 180, 52–69 (2015).

[3] A. S. Fraenkel and J. Simpson: How many squares can a string contain? *J. Comb. Theory, Ser. A*, 82(1), 112–120 (1998).

[4] L. Ilie: A note on the number of squares in a word. *Theor. Comput. Sci.*, 380(3), 373–376 (2007).

[5] M. Kubica, J. Radoszewski, W. Rytter, and T. Waleń: On the maximum number of cubic subwords in a word. *Eur. J. Comb.* 34(1), 27–37 (2013).

[6] N. H. Lam: On the number of squares in a string. *AdvOL-Report* 2 (2013).

[7] S. Li: On the number of k-powers in a finite word. *Adv. Appl. Math.* 139, 102371 (2022).

[8] M. Lothaire: Combinatorics on Words. Cambridge University Press, Cambridge (1997).

[9] M. Lothaire: Applied Combinatorics on Words. Cambridge University Press, Cambridge (2005).

[10] A. Thierry: A proof that a word of length n has less than 1.5n distinct squares (2020), arXiv:2001.02996.