Asymptotic bounds for the number of closed and privileged words

Daniel Gabric
Department of Math/Stats
University of Winnipeg
Winnipeg, MB R3B 2E9
Canada
d.gabric@uwinnipeg.ca

Abstract

A word w has a border u if u is a non-empty proper prefix and suffix of u. A word w is said to be *closed* if w is of length at most 1 or if w has a border that occurs exactly twice in w. A word w is said to be *privileged* if w is of length at most 1 or if w has a privileged border that occurs exactly twice in w. Let $C_k(n)$ (resp. $P_k(n)$) be the number of length-n closed (resp. privileged) words over a k-letter alphabet. In this paper, we improve existing upper and lower bounds on $C_k(n)$ and $P_k(n)$. We prove that $C_k(n) \in \Theta(\frac{k^n}{n})$. Let $\log_k^{\circ 0}(n) = n$ and $\log_k^{\circ j}(n) = \log_k(\log_k^{\circ j-1}(n))$ for $j \geq 1$. We also prove that for all $j \geq 0$ there exist constants N_j, c_j, and c'_j such that

$$c_j \frac{k^n}{n \log_k^{\circ j}(n) \prod_{i=1}^{j} \log_k^{\circ i}(n)} \leq P_k(n) \leq c'_j \frac{k^n}{n \prod_{i=1}^{j} \log_k^{\circ i}(n)}$$

for all $n > N_j$.

1 Introduction

Let Σ_k denote the k-letter alphabet $\{0, 1, \ldots, k-1\}$. Throughout this paper, we denote the length of a word w as $|w|$. A word u is said to be a *factor* of a word w if $w = xuy$ for some words x, y. A word w has a *border* u if u is a non-empty proper prefix and suffix of w. A word that has a border is said to be *bordered*; otherwise, it is said to be *unbordered*. A word w is said to be *closed* if $|w| \leq 1$ or if w has a border that occurs exactly twice in w. If u is a border w and u occurs in w exactly twice, then we say w is *closed by* u. It is easy to see that if a word w is closed by a word u, then u must be the largest border in w; otherwise u would occur more than two times in w. A word w is said to be *privileged* if $|w| \leq 1$ or if w is closed by a privileged word.

Example 1. The English word `entanglement` has the border `ent` and only contains two occurrences of `ent`. Thus, `entanglement` is a closed word, closed by `ent`. Since $|\texttt{ent}| > 1$ and `ent` is unbordered and therefore not privileged, we have that `entanglement` is not privileged.

The English word `alfalfa` is closed by `alfa`. Furthermore, `alfa` is closed by `a`. But $|\texttt{a}| \leq 1$, so `alfa` is privileged and therefore so is `alfalfa`.

The only border of the English word `eerie` is `e` and `e` appears 3 times in the word. Thus, `eerie` is neither closed nor privileged.

Closed words were introduced relatively recently by Fici [5] as a way to classify Trapezoidal and Sturmian words. However, there are multiple equivalent formulations of closed words that have been defined at different times. Closed words are equivalent to codewords in prefix-synchronized codes [8, 9]. Closed words are also equivalent to *periodic-like* words [3]. A *period* of a word $w = w_1 w_2 \cdots w_n$ is an integer $p \leq n$ such that $w_i = w_{i+p}$ for all $1 \leq i \leq n - p$. A length-n word is said to be *periodic* if it has a period of length $\leq n/2$. In applications that require the analysis of long words, like DNA sequence analysis, the smallest period is typically much larger than half the length of the word. Periodic-like words were introduced as a generalization of periodic words that preserve some desirable properties of periodic words.

Privileged words [13] were introduced as a technical tool related to a problem in dynamical systems and discrete geometry. They were originally defined as a generalization of *rich* words by tweaking the definition of a *complete first return*. A word is said to be *rich* if it contains, as factors, the maximum possible number of distinct palindromes. A *palindrome* is a word that reads the same forwards and backwards. A *complete first return* to a word u is a word that starts and ends with u, and contains only two occurrences of u. An equivalent definition of rich words is the following: a word w is said to be rich if and only if every palindromic factor of w is a complete first return to a shorter palindrome. Privileged words were then defined as an iterated complete first return. A word is privileged if and only if it is a complete first return to a shorter privileged word. Single letters and the empty word are defined to be privileged in order to make this definition meaningful.

Since their introduction, there has been much research into the properties of closed and privileged words [1, 2, 4, 6, 12, 16, 17, 20]. One problem that has received some interest lately [7, 14, 18, 19] is to find good upper and lower bounds for the number of closed and privileged words.

Let $C_k(n)$ denote the number of length-n closed words over Σ_k. Let $C_k(n,t)$ denote the number of length-n closed words over Σ_k that are closed by a length-t word. Let $P_k(n)$ denote the number of length-n privileged words over Σ_k. Let $P_k(n,t)$ denote the number of length-n privileged words over Σ_k that are closed by a length-t privileged word. See Tables 1 and 2 for sample values of $C_2(n,t)$ and $P_2(n,t)$ for small n, t. See sequences A226452 and A231208 in the *On-Line Encyclopedia of Integer Sequences* [15] for sample values of $C_2(n)$ and $P_2(n)$.

Every privileged word is a closed word, so any upper bound on $C_k(n)$ is also an upper bound on $P_k(n)$. Furthermore, any lower bound on $P_k(n)$ is also a lower bound on $C_k(n)$.

- Forsyth et al. [7] showed that $P_2(n) \geq 2^{n-5}/n^2$ for all $n \geq n_0$ for some $n_0 > 0$.

- Nicholson and Rampersad [14] improved and generalized this bound by showing that there are constants c and n_0 such that $P_k(n) \geq c\frac{k^n}{n(\log_k(n))^2}$ for all $n \geq n_0$.

- Rukavicka [18] showed that there is a constant c such that $C_k(n) \leq c \ln n \frac{k^n}{\sqrt{n}}$ for all $n > 1$.

- Rukavicka [19] also showed that for every $j \geq 3$, there exist constants α_j and n_j such that $P_k(n) \leq \alpha_j \frac{k^n\sqrt{\ln n}}{\sqrt{n}} \ln^{\circ j}(n) \prod_{i=2}^{j-1} \sqrt{\ln^{\circ i}(n)}$ length-n privileged words for all $n \geq n_j$ where $\ln^{\circ 0}(n) = n$ and $\ln^{\circ j}(n) = \ln(\ln^{\circ j-1}(n))$.

The best upper and lower bounds for both $C_k(n)$ and $P_k(n)$ are widely separated, and can be much improved. In this paper, we improve the existing upper and lower bounds on $C_k(n)$ and $P_k(n)$. We prove the following two theorems.

Theorem 2. *Let $k \geq 2$ be an integer.*

(a) There exist constants N and c such that $C_k(n) \geq c\frac{k^n}{n}$ for all $n > N$.

(b) There exist constants N' and c' such that $C_k(n) \leq c'\frac{k^n}{n}$ for all $n > N'$.

Theorem 3. *Let $k \geq 2$ be an integer. Let $\log_k^{\circ 0}(n) = n$ and $\log_k^{\circ j}(n) = \log_k(\log_k^{\circ j-1}(n))$ for $j \geq 1$.*

(a) For all $j \geq 0$ there exist constants N_j and c_j such that

$$P_k(n) \geq c_j \frac{k^n}{n \log_k^{\circ j}(n) \prod_{i=1}^{j} \log_k^{\circ i}(n)}$$

for all $n > N_j$.

(b) For all $j \geq 0$ there exist constants N'_j and c'_j such that

$$P_k(n) \leq c'_j \frac{k^n}{n \prod_{i=1}^{j} \log_k^{\circ i}(n)}$$

for all $n > N'_j$.

Before we proceed, we give a heuristic argument as to why $C_k(n)$ is in $\Theta(\frac{k^n}{n})$. Consider a "random" length-n word w. Let $\ell = \log_k(n) + c$ where c is a constant such that ℓ is a positive integer. There is a $\frac{1}{k^\ell} = \frac{1}{k^c n}$ chance that w has a length-ℓ border. Suppose w has a length-ℓ border. Now suppose we drop the first and last character of w to get w'. If w' were randomly chosen (which it is not), then we could use the linearity of expectation to get that the expected number of occurrences of u in w' is approximately $(n - 1 - \ell)k^{-\ell} \approx k^{-c}$. Thus, for c large enough we have that u does not occur in w' with high probability, and so w is closed. Therefore, there are approximately $k^{n-\ell} \in \Theta(\frac{k^n}{n})$ length-n closed words.

n \ t	1	2	3	4	5	6	7	8	9	10
10	2	30	70	50	30	12	6	2	2	0
11	2	42	118	96	54	30	13	6	2	2
12	2	60	200	182	114	54	30	12	6	2
13	2	88	338	346	214	126	54	30	12	6
14	2	132	570	640	432	232	126	54	30	12
15	2	202	962	1192	828	474	240	126	54	30
16	2	314	1626	2220	1612	908	492	240	126	54
17	2	494	2754	4128	3112	1822	956	504	240	126
18	2	784	4676	7670	6024	3596	1934	982	504	240
19	2	1252	7960	14264	11636	7084	3828	1992	990	504
20	2	2008	13588	26524	22512	13928	7632	3946	2026	990

Table 1: Some values of $C_2(n,t)$ for n, t where $10 \le n \le 20$ and $1 \le t \le 10$.

n \ t	1	2	3	4	5	6	7	8	9	10
10	2	16	22	8	6	2	2	0	2	0
11	2	26	38	16	10	6	4	2	2	2
12	2	42	68	30	18	4	6	2	2	0
13	2	68	122	58	38	14	10	6	4	2
14	2	110	218	108	76	20	14	8	6	2
15	2	178	390	204	148	46	24	18	14	6
16	2	288	698	384	288	86	48	16	18	8
17	2	466	1250	724	556	178	92	36	32	26
18	2	754	2240	1364	1076	344	190	64	36	28
19	2	1220	4016	2572	2092	688	388	136	70	56
20	2	1974	7204	4850	4068	1342	772	268	138	52

Table 2: Some values of $P_2(n,t)$ for n, t where $10 \le n \le 20$ and $1 \le t \le 10$.

2 Preliminary results

In this section we give some necessary results and definitions in order to prove our main results. Also throughout this paper, we use c's, d's, and N's to denote positive real constants (dependent on k).

Let w be a length-n word. Suppose w is closed by a length-t word u. Since u is also the largest border of w, it follows that w cannot be closed by another word. This implies that

$$C_k(n) = \sum_{i=1}^{n-1} C_k(n,t) \text{ and } P_k(n) = \sum_{i=1}^{n-1} P_k(n,t)$$

for $n > 1$.

Let $B_k(n,u)$ denote the number of length-n words over Σ_k that are closed by the word u. Let $A_k(n,u)$ denote the number of length-n words over Σ_k that do not contain the word u as a factor.

The *auto-correlation* [9, 10, 11] of a length-t word u is a length-t binary word $a(u) = a_1a_2\cdots a_t$ where $a_i = 1$ if and only if u has a border of length $t-i+1$. The *auto-correlation polynomial* $f_{a(u)}(z)$ of $a(u)$ is defined as

$$f_{a(u)}(z) = \sum_{i=0}^{t-1} a_{t-i}z^i.$$

For example, the word u = `entente` has auto-correlation $a(u) = 1001001$ and auto-correlation polynomial $f_{a(u)}(z) = z^6 + z^3 + 1$.

We now prove two technical lemmas that will be used in the proofs of Theorem 2 (b) and Theorem 3 (b).

Lemma 4. *Let $k, t \geq 2$ be integers, and let γ be a real number such that $0 < \gamma \leq \frac{6}{t}$. Then*

$$k^t - \gamma t k^{t-1} \leq (k-\gamma)^t \leq k^t - \gamma t k^{t-1} + \frac{1}{2}\gamma^2 t(t-1)k^{t-2}.$$

Proof. The case when $k = 2$ was proved in a paper by Forsyth et al. [7, Lemma 9]. We generalize their proof to $k \geq 3$.

When $t = 2$, we have $k^2 - 2k\gamma \leq (k-\gamma)^2 \leq k^2 - 2k\gamma + \gamma^2$. So suppose $t \geq 3$. By the binomial theorem, we have

$$\begin{aligned}(k-\gamma)^t = \sum_{i=0}^{t} k^{t-i}(-\gamma)^i\binom{t}{i} &= k^t - \gamma t k^{t-1} + \sum_{i=2}^{t} k^{t-i}(-\gamma)^i\binom{t}{i}\\ &\geq k^t - \gamma t k^{t-1} + \sum_{j=1}^{\lfloor (t-1)/2\rfloor}\left(k^{t-2j}\gamma^{2j}\binom{t}{2j} - k^{t-2j-1}\gamma^{2j+1}\binom{t}{2j+1}\right).\end{aligned}$$

So to show that $k^t - \gamma t k^{t-1} \leq (k-\gamma)^t$, it is sufficient to show that

$$k^{t-2j}\gamma^{2j}\binom{t}{2j} \geq k^{t-2j-1}\gamma^{2j+1}\binom{t}{2j+1} \tag{1}$$

for $1 \leq j \leq \lfloor (t-1)/2 \rfloor \leq (t-1)/2$.

By assumption we have that $\gamma \leq \frac{6}{t}$, so $\gamma \leq \frac{6}{t-2}$ and thus $\gamma t - 2\gamma \leq 6$. Adding $2\gamma - 2$ to both sides we get $\gamma t - 2 \leq 4 + 2\gamma$, and so $\frac{\gamma t-2}{\gamma+2} \leq 2$. If $i \geq 2 \geq \frac{\gamma t - 2}{\gamma+2}$, then $(\gamma+2)i \geq \gamma t - 2$. This implies that $2(i+1) \geq \gamma(t-i)$, and

$$\frac{k}{\gamma} \geq \frac{2}{\gamma} \geq \frac{t-i}{i+1} = \frac{\binom{t}{i+1}}{\binom{t}{i}}.$$

Therefore letting $i = 2j$, we have that $k\binom{t}{2j} \geq \gamma\binom{t}{2j+1}$. Multiplying both sides by $k^{t-2j-1}\gamma^{2j}$ we get $k^{t-2j}\gamma^{2j}\binom{t}{2j} \geq k^{t-2j-1}\gamma^{2j+1}\binom{t}{2j+1}$, which proves (1).

Now we prove that $(k-\gamma)^t \leq k^t - \gamma t k^{t-1} + \frac{1}{2}\gamma^2 t(t-1)k^{t-2}$. Going back to the binomial expansion of $(k-\gamma)^t$, we have

$$\begin{aligned}(k-\gamma)^t &= k^t - \gamma t k^{t-1} + \frac{1}{2}\gamma^2 t(t-1)k^{t-2} + \sum_{i=3}^{t} k^{t-i}(-\gamma)^i \binom{t}{i} \\ &\leq k^t - \gamma t k^{t-1} + \frac{1}{2}\gamma^2 t(t-1)k^{t-2} \\ &\quad - \sum_{j=1}^{\lfloor (t-2)/2 \rfloor} \left(k^{t-2j-1}\gamma^{2j+1}\binom{t}{2j+1} - k^{t-2j-2}\gamma^{2j+2}\binom{t}{2j+2} \right).\end{aligned}$$

So to show that $(k-\gamma)^t \leq k^t - \gamma t k^{t-1} + \frac{1}{2}\gamma^2 t(t-1)k^{t-2}$, it is sufficient to show that

$$k^{t-2j-1}\gamma^{2j+1}\binom{t}{2j+1} \geq k^{t-2j-2}\gamma^{2j+2}\binom{t}{2j+1}$$

for $1 \leq j \leq \lfloor (t-2)/2 \rfloor$. But we have already proved that $k\binom{t}{i} \geq \gamma\binom{t}{i+1}$. Letting $i = 2j$, we have that $k\binom{t}{2j+1} \geq \gamma\binom{t}{2j+2}$. Multiplying both sides by $k^{t-2j-2}\gamma^{2j+1}$ we get $k^{t-2j-1}\gamma^{2j+1}\binom{t}{2j+1} \geq k^{t-2j-2}\gamma^{2j+2}\binom{t}{2j+2}$.

□

Let $\log_k^{\circ 0}(n) = n$ and $\log_k^{\circ j}(n) = \log_k(\log_k^{\circ j-1}(n))$ for $j \geq 1$.

Lemma 5. *Let $i \geq 1$ and $k \geq 2$ be integers. Then for any constant $\gamma > 0$, we have*

$$\lim_{n\to\infty} \frac{\log_k^{\circ i}(n^\gamma)}{\log_k^{\circ i}(n)} = \begin{cases} \gamma, & \text{if } i = 1; \\ 1, & \text{if } i > 1. \end{cases}.$$

Proof. When $i = 1$ we have $\lim_{n\to\infty} \frac{\log_k(n^\gamma)}{\log_k(n)} = \gamma \lim_{n\to\infty} \frac{\log_k(n)}{\log_k(n)} = \gamma$.

The proof is by induction on i. Since we will use L'Hôpital's rule to evaluate the limit, we first compute the derivative of $\log_k^{\circ i}(n^\lambda)$ with respect to n for any constant $\lambda > 0$. We have

$$\frac{d}{dn} \log_k^{\circ i}(n^\lambda) = \frac{\lambda}{n \prod_{j=1}^{i-1} \log_k^{\circ j}(n^\lambda)}.$$

In the base case, when $i = 2$, we have

$$\lim_{n\to\infty} \frac{\log_k^{\circ 2}(n^\gamma)}{\log_k^{\circ 2}(n)} = \lim_{n\to\infty} \frac{\frac{\gamma}{n \log_k(n^\gamma)}}{\frac{1}{n \log_k(n)}} = 1.$$

Suppose $i > 2$. Then we have

$$\lim_{n\to\infty} \frac{\log_k^{\circ i}(n^\gamma)}{\log_k^{\circ i}(n)} = \lim_{n\to\infty} \frac{\frac{\gamma}{n\prod_{j=1}^{i-1}\log_k^{\circ j}(n^\gamma)}}{\frac{1}{n\prod_{j=1}^{i-1}\log_k^{\circ j}(n)}} = \lim_{n\to\infty} \frac{\prod_{j=2}^{i-1}\log_k^{\circ j}(n)}{\prod_{j=2}^{i-1}\log_k^{\circ j}(n^\gamma)} = 1.$$

□

3 Closed words

3.1 Lower bound

We first state a useful lemma from a paper of Nicholson and Rampersad [14].

Lemma 6 (Nicholson and Rampersad [14]). *Let $k \geq 2$ be an integer. For every n, there is a unique integer t such that*

$$\frac{\ln k}{k-1} k^t \leq n - t < \frac{\ln k}{k-1} k^{t+1}.$$

Let u be a length-t word. There exist constants N_0 and d such that for $n - t > N_0$ we have

$$B_k(n,u) \geq d\frac{k^n}{n^2}.$$

We now use the previous lemma to prove Theorem 2 (a).

Proof of Theorem 2 (a). The number $C_k(n,t)$ of length-n words closed by a length-t word is clearly equal to the sum, over all length-t words u, of the number $B_k(n,u)$ of length-n words closed by u. Thus we have that

$$C_k(n,t) = \sum_{|u|=t} B_k(n,u).$$

Let $t = \lfloor \log_k(n-t) + \log_k(k-1) - \log_k(\ln k) \rfloor$. By Lemma 6 there exist constants N_0 and d such that for $n - t > N_0$ we have $B_k(n,u) \geq dk^n/n^2$. Clearly $t \leq \log_k(n) + 1$ for all $n \geq 1$. Since t is asymptotically much smaller than n, there exists a constant $N > N_0$ such that $n - t > N_0$ for all $n > N$. Thus for $n > N$ we have

$$\begin{aligned} C_k(n) \geq C_k(n,t) &= \sum_{|u|=t} B_k(n,u) \geq \sum_{|u|=t} d\frac{k^n}{n^2} = k^t\left(d\frac{k^n}{n^2}\right) \\ &= dk^{\lfloor \log_k(n-t)+\log_k(k-1)-\log_k(\ln k)\rfloor}\frac{k^n}{n^2} \geq d_0 k^{\log_k(n-t)+\log_k(k-1)-\log_k(\ln k)}\frac{k^n}{n^2} \\ &\geq d_1(n-t)\frac{k^n}{n^2} \geq d_1(n - \log_k(n) - 1)\frac{k^n}{n^2} \geq c\frac{k^n}{n} \end{aligned}$$

for some constant $c > 0$. □

3.2 Upper bound

Before we proceed with upper bounding $C_k(n)$, we briefly outline the direction of the proof. First, we begin by bounding $C_k(n,t)$ for $t < n/2$ and $t \geq n/2$. We show that for $t < n/2$, the number of length-n words closed by a particular length-t word u is bounded by the number of words of length $n-2t$ that do not have 0^t as a factor. For $t \geq n/2$ we prove that $C_k(n,t)$ is negligibly small. Next, we prove upper bounds on the number of words that do not have 0^t as a factor, allowing us to finally bound $C_k(n)$.

Lemma 7. *Let n, t, and k be integers such that $n \geq 2t \geq 2$ and $k \geq 2$. Let u be a length-t word. Then*

$$B_k(n,u) \leq A_k(n-2t, 0^t).$$

Proof. Recall that $B_k(n,u)$ is the number of length-n words that are closed by the word u. Also recall that $A_k(n,u)$ is the number of length-n words that do not contain the word u as a factor.

Let w be a length-n word closed by u where $|w| = n \geq 2t = 2|u|$. Then we can write $w = uvu$ where v does not contain u as a factor. This immediately implies that $B_k(n,u) \leq A_k(n-2t,u)$. But from a result of Guibas and Odlyzko [11, Section 7], we have that if $f_{a(u)}(2) > f_{a(v)}(2)$ for words u, v, then $A_k(m,u) \geq A_k(m,v)$ for all $m \geq 1$. The auto-correlation polynomial only has 0 or 1 as coefficients, depending on the 1's and 0's in the auto-correlation. Thus, the auto-correlation p that maximizes $f_p(2)$ is clearly $p = 1^t$. The words that achieve this auto-correlation are words of the form a^t where $a \in \Sigma_k$. Therefore we have

$$B_k(n,u) \leq A_k(n-2t,u) \leq A_k(n-2t,0^t).$$

□

Lemma 8. *Let n, t, and k be integers such that $n \geq 2t \geq 2$ and $k \geq 2$. Then*

$$C_k(n,t) \leq k^t A_k(n-2t,0^t).$$

Proof. The number $C_k(n,t)$ of length-n words closed by a length-t word is equal to the sum, over all length-t words u, of the number $B_k(n,u)$ of length-n words closed by u. Thus we have that

$$C_k(n,t) = \sum_{|u|=t} B_k(n,u).$$

By Lemma 7 we have that $B_k(n,v) \leq A_k(n-2t,0^t)$ for all length-t words v. Therefore

$$C_k(n,t) = \sum_{|u|=t} B_k(n,u) \leq \sum_{|u|=t} A_k(n-2t,0^t) \leq k^t A_k(n-2t,0^t).$$

□

Corollary 9. *Let $n \geq 1$ and $k \geq 2$ integers. Then*

$$C_k(n) \leq \sum_{t=1}^{\lfloor n/2 \rfloor} k^t A_k(n-2t,0^t) + nk^{\lceil n/2 \rceil}.$$

Proof. It follows from Lemma 8 that

$$C_k(n) = \sum_{t=1}^{n-1} C_k(n,t) \leq \sum_{t=1}^{\lfloor n/2 \rfloor} k^t A_k(n-2t, 0^t) + \sum_{t=\lfloor n/2 \rfloor + 1}^{n-1} C_k(n,t).$$

Now we show that

$$\sum_{t=\lfloor n/2 \rfloor + 1}^{n-1} C_k(n,t) \leq nk^{\lceil n/2 \rceil}.$$

Let $w = w_0 w_1 \cdots w_{n-1}$ be a word of length n that is closed by a word u of length $t > \lfloor n/2 \rfloor$. Then $w = ux = yu$ for some words x, y. So $w_i = w_{i+(n-t)}$ for all i, $0 \leq i < t$. This implies that $w = v^i v'$ where v is the length-$(n-t)$ prefix of w, $i = \lfloor n/|v| \rfloor$, and v' is the length-$(n - i|v|)$ prefix of v. Since $t > \lfloor n/2 \rfloor$, we have that $n - t < \lceil n/2 \rceil$. We see that w is fully determined by the word v. So since $|v| < \lceil n/2 \rceil$, we have $C_k(n,t) \leq k^{\lceil n/2 \rceil}$. Thus

$$\sum_{t=\lfloor n/2 \rfloor + 1}^{n-1} C_k(n,t) \leq \sum_{t=\lfloor n/2 \rfloor + 1}^{n-1} k^{\lceil n/2 \rceil} \leq nk^{\lceil n/2 \rceil}.$$

□

Lemma 10. *Let $n \geq 0$, $t \geq 1$, and $k \geq 2$ be integers. Then*

$$A_k(n, 0^t) = \begin{cases} k^n, & \text{if } n < t; \\ (k-1) \sum_{i=1}^{t} A_k(n-i, 0^t), & \text{if } n \geq t. \end{cases}$$

Proof. If $n < t$, then any length-n word is shorter than 0^t, and thus cannot contain 0^t as a factor. So $A_k(n, 0^t) = k^n$.

Suppose $n \geq t$. Let w be a length-n word that does not contain 0^t as a factor. Let us look at the symbols that w ends in. Since w does not contain 0^t, we have that w ends in anywhere from 0 to $t-1$ zeroes. So w is of the form $w = w' b 0^i$ where i is an integer with $0 \leq i \leq t-1$, $b \in \Sigma_k - \{0\}$, and w' is a length-$(n-i-1)$ word that does not contain 0^t as a factor. There are $k-1$ choices for b, and $A_k(n-i-1, 0^t)$ choices for w'. So there are $(k-1)A_k(n-i-1, 0^t)$ words of the form $w'b0^i$. Summing over all possible i gives

$$A_k(n, 0^t) = (k-1) \sum_{i=1}^{t} A_k(n-i, 0^t).$$

□

Corollary 11. *Let $n \geq 0$, $t \geq 1$, and $k \geq 2$ be integers. Then*

$$A_k(n, 0^t) = \begin{cases} k^n, & \text{if } n < t; \\ k^n - 1, & \text{if } n = t; \\ kA_k(n-1, 0^t) - (k-1)A_k(n-t-1, 0^t), & \text{if } n > t. \end{cases}$$

Proof. Compute $A_k(n, 0^t) - A_k(n-1, 0^t)$ with the recurrence from Lemma 10 and the result follows. □

Corollary 12. *Let $n \geq 0$, $t \geq 1$, and $k \geq 2$ be integers. Then*

$$A_k(n, 0^t) = \begin{cases} k^n, & \text{if } n < t; \\ k^{n-t}(k^t - 1) - (n-t)k^{n-t-1}(k-1), & \text{if } t \leq n \leq 2t; \\ kA_k(n-1, 0^t) - (k-1)A_k(n-t-1, 0^t), & \text{if } n > 2t. \end{cases}$$

Proof. We prove $t \leq n \leq 2t$ by induction on n. In the base case, when $n = t$, we have $k^t - 1 = A_k(t, 0^t) = k^{t-t}(k^t - 1) - (t-t)k^{t-t-1}(k-1) = k^t - 1$.

Suppose $t < n \leq 2t$. Then

$$\begin{aligned} A_k(n, 0^t) &= kA_k(n-1, 0^t) - (k-1)A_k(n-t-1, 0^t) \\ &= k(k^{n-1-t}(k^t - 1) - (n-1-t)k^{n-t-2}(k-1)) - (k-1)k^{n-t-1} \\ &= k^{n-t}(k^t - 1) - (n-t)k^{n-t-1}(k-1). \end{aligned}$$

□

Since $(A_k(n, 0^t))_n$ satisfies a linear recurrence, we know that the asymptotic behaviour of $A_k(n, 0^t)$ is determined by the root of maximum modulus of the polynomial $x^{t+1} - kx^t + k - 1 = 0$. We use this fact to find an upper bound for $A_k(n, 0^t)$.

Lemma 13. *Let $t \geq 1$ and $k \geq 2$ be integers. Let*

$$\beta_k(t) = k - (k-1)k^{-t-1}.$$

Then $\beta_k(t) \geq k - (k-1)\beta_k(t)^{-t}$.

Proof. Since $\beta_k(t) \leq k$, we have that $\beta_k(t)^{-t} \geq k^{-t} \geq k^{-t-1}$. This implies that

$$\beta_k(t) = k - (k-1)k^{-t-1} \geq k - (k-1)\beta_k(t)^{-t}.$$

□

Lemma 14. *Let $k, t \geq 2$ be integers. Let n be an integer such that $2t \leq n \leq 3t$. Then $A_k(n, 0^t) \leq \beta_k(t)^n$.*

Proof. The proof is by induction on n. By Corollary 12 we have that $A_k(n, 0^t) = k^{n-t}(k^t - 1) - (n-t)k^{n-t-1}(k-1)$ for $t \leq n \leq 2t$. Let $\gamma(t) = (k-1)k^{-t-1}$.

Suppose, for the base case, that $n = 2t$. Then

$$\begin{aligned} A_k(2t, 0^t) &= k^t(k^t - 1) - tk^{t-1}(k-1) = k^{2t} - k^{t-2}(k^2 + tk(k-1)) \\ &= k^{2t} - \gamma(t)k^{2t-1}\frac{(k^2 + tk(k-1))}{k-1} \\ &\leq k^{2t} - \gamma(t)tk^{2t-1}. \end{aligned}$$

Clearly $\gamma(t) \le 6/t$ for all $t \ge 2$, so $A_k(2t) \le k^{2t} - \gamma(t)tk^{2t-1} \le (k-\gamma(t))^{2t} = \beta_k(t)^{2t}$.

Suppose that $2t < n \le 3t$. Furthermore let $n = 2t+i+1$ where i is an integer such that $0 \le i < t$. Notice that $A_k(n-t-1, 0^t) = A_k(t+i, 0^t) = k^i(k^t-1) - ik^{i-1}(k-1)$. Then

$$\begin{aligned} A_k(2t+i+1, 0^t) &= kA_k(2t+i, 0^t) - (k-1)A_k(t+i, 0^t) \\ &\le k(k-\gamma(t))^{2t+i} - (k-1)(k^i(k^t-1) - ik^{i-1}(k-1)) \\ &= (k-\gamma(t))^{2t+i+1} + \gamma(t)(k-\gamma(t))^{2t+i} - (k-1)(k^i(k^t-1) - ik^{i-1}(k-1)) \\ &= \beta_k(t)^{2t+i+1} + \gamma(t)\beta_k(t)^{2t+i} - (k-1)(k^i(k^t-1) - ik^{i-1}(k-1)). \end{aligned}$$

To prove the desired bound, namely that $A_k(2t+i+1, 0^t) \le \beta_k(t)^{2t+i+1}$, it is sufficient to show that $\beta_k(t)^{2t+i} \le \gamma(t)^{-1}(k-1)(k^i(k^t-1) - ik^{i-1}(k-1))$. We begin by upper bounding $\beta_k(t)^{2t+i}$ with Lemma 4. We have

$$\begin{aligned} \beta_k(t)^{2t+i} &\le k^{2t+i} - \gamma(t)(2t+i)k^{2t+i-1} + \frac{1}{2}\gamma(t)^2(2t+i)(2t+i-1)k^{2t+i-2} \\ &\le k^{2t+i} - 2(k-1)tk^{t+i-2} + \frac{9}{2}(k-1)^2t^2k^{i-4} \\ &\le k^{2t+i+1} - (k-1)k^{2t+i} - 2(k-1)tk^{t+i-2} + \frac{9}{2}(k-1)^2t^2k^{i-4} \\ &= k^{2t+i+1} - k^{t+i}\Big((k-1)k^t + 2(k-1)tk^{-2} - \frac{9}{2}(k-1)^2t^2k^{-t-4}\Big). \end{aligned} \tag{2}$$

It is easy to verify that $(k-1)k^t \ge k + t(k-1)$ and $2(k-1)tk^{-2} - \frac{9}{2}(k-1)^2t^2k^{-t-4} \ge 0$ for all $t \ge 2$. Thus, continuing from (2), we have

$$\begin{aligned} \beta_k(t)^{2t+i} &\le k^{2t+i+1} - k^{t+i}(k+t(k-1)) \le k^{2t+i+1} - k^{t+i}(k+i(k-1)) \\ &= \frac{k^{t+1}}{k-1}(k-1)(k^{t+i} - k^i - ik^{i-1}(k-1)) \\ &= \gamma(t)^{-1}(k-1)(k^i(k^t-1) - ik^{i-1}(k-1)). \end{aligned}$$

□

Lemma 15. *Let n, t, and k be integers such that $n \ge 2t \ge 4$ and $k \ge 2$. Then $A_k(n, 0^t) \le \beta_k(t)^n$.*

Proof. The proof is by induction on n. The base case, when $2t \le n \le 3t$, is taken care of in Lemma 14.

Suppose $n > 3t$. Then

$$A_k(n, 0^t) = (k-1)\sum_{i=1}^{t} A_k(n-i, 0^t) \le (k-1)\sum_{i=1}^{t} \beta_k(t)^{n-i} = (k-1)\frac{\beta_k(t)^n - \beta_k(t)^{n-t}}{\beta_k(t)-1}.$$

By Theorem 13, we have that $\beta_k(t) - 1 \ge (k-1) - (k-1)\beta_k(t)^{-t}$. Therefore

$$A_k(n, 0^t) \le (k-1)\frac{\beta_k(t)^n - \beta_k(t)^{n-t}}{\beta_k(t)-1} = \beta_k(t)^n \frac{(k-1) - (k-1)\beta_k(t)^{-t}}{\beta_k(t)-1} \le \beta_k(t)^n.$$

□

Proof of Theorem 2 (b). First notice that $A_k(n,0) = (k-1)^n$, since $A_k(n,0)$ is just the number of length-n words that do not contain 0. Bounds for the number of closed and privileged words

Let N' be a positive integer such that the following inequalities hold for all $n > N'$.

$$\begin{aligned}
C_k(n) &\le \sum_{t=2}^{\lfloor n/2\rfloor} k^t A_k(n-2t, 0^t) + kA_k(n-2,0) + nk^{\lceil n/2\rceil} \\
&\le \sum_{t=2}^{\lfloor n/2\rfloor} k^t \beta_k(t)^{n-2t} + k(k-1)^{n-2} + nk^{\lceil n/2\rceil} \\
&\le \sum_{t=2}^{\lfloor n/2\rfloor} k^t \left(k - \frac{k-1}{k^{t+1}}\right)^{n-2t} + d_2\frac{k^n}{n} = k^n \sum_{t=2}^{\lfloor n/2\rfloor} \frac{1}{k^t}\left(1 - \frac{k-1}{k^{t+2}}\right)^{n-2t} + d_2\frac{k^n}{n} \\
&\le k^n \left(\sum_{t=2}^{\lfloor \log_k n\rfloor} \frac{1}{k^t}\left(1 - \frac{k-1}{k^{t+2}}\right)^{n-2t} + \sum_{t=\lfloor \log_k n\rfloor+1}^{\lfloor n/2\rfloor} \frac{1}{k^t}\right) + d_2\frac{k^n}{n} \\
&\le k^n \left(\sum_{t=2}^{\lfloor \log_k n\rfloor} \frac{1}{k^t}\left(1 - \frac{k-1}{k^{t+2}}\right)^{n-2\lfloor \log_k n\rfloor} + \frac{d_3}{n}\right) + d_2\frac{k^n}{n} \\
&\le k^n \sum_{t=2}^{\lfloor \log_k n\rfloor} \frac{1}{k^t}\left(1 - \frac{k-1}{k^{t+2}}\right)^{n/2} + d_4\frac{k^n}{n}.
\end{aligned} \tag{3}$$

Now we bound the sum in (3). Let $h(x) = (1-(k-1)k^{-2}x)^{n/2}$. Notice that $h(x)$ is monotonically decreasing on the interval $x \in (0,1)$. So for $k^{-t-1} \le x \le k^{-t}$ we have that $h(x) \ge h(k^{-t})$. Thus

$$\frac{1}{k^t}\left(1 - \frac{k-1}{k^{t+2}}\right)^{n/2} \le \frac{k-1}{k^t}\left(1 - \frac{k-1}{k^{t+2}}\right)^{n/2} \le k\left(\left(\frac{1}{k^t} - \frac{1}{k^{t+1}}\right)h(k^{-t})\right) \le k\int_{k^{-t-1}}^{k^{-t}} h(x)\,dx.$$

Going back to (3) we have

$$C_k(n) \le k^n \sum_{t=2}^{\lfloor \log_k n\rfloor} k\int_{k^{-t-1}}^{k^{-t}} h(x)\,dx + d_4\frac{k^n}{n} \le k^{n+1}\int_0^1 h(x)\,dx + d_4\frac{k^n}{n}.$$

Evaluating and bounding the definite integral, we have

$$\begin{aligned}
\int_0^1 h(x)\,dx &= -\frac{k^2}{k-1}\left[\frac{(1-(k-1)k^{-2}x)^{n/2+1}}{n/2+1}\right]_{x=0}^{x=1} \\
&= -\frac{k^2}{k-1}\left(\frac{(1-(k-1)k^{-2})^{n/2+1} - 1}{n/2+1}\right) \\
&\le d_5\left(\frac{1-(1-(k-1)k^{-2})^{n/2+1}}{n/2+1}\right) \le d_5\frac{1}{n/2+1} \le \frac{d_6}{n}.
\end{aligned}$$

Putting everything together, we have that

$$C_k(n) \le k^{n+1} \int_0^1 h(x)\,dx + d_4 \frac{k^n}{n} \le d_6 \frac{k^{n+1}}{n} + d_4 \frac{k^n}{n} \le c' \frac{k^n}{n}$$

for some constant $c' > 0$. □

4 Privileged words

4.1 Lower bound

In this section we provide a family of lower bounds for the number of length-n privileged words. We use induction to prove these bounds. The basic idea is that we start with the lower bound by Nicholson and Rampersad, and then use it to bootstrap ourselves to better and better lower bounds.

Proof of Theorem 3 (a). The proof is by induction on j. Let $t = \lfloor \log_k(n-t) + \log_k(k-1) - \log_k(\ln k) \rfloor$. We clearly have $0 \le t \le \log_k(n) + 1$ for all $n \ge 1$. Let u be a length-t privileged word. By Lemma 6 we have that there exist constants N_0 and c_0 such that $P_k(n) \ge B_k(n,u) \ge c_0 \frac{k^n}{n^2}$ for all $n > N_0$. So the base case, when $j = 0$, is taken care of.

Suppose $j > 0$. By induction we have that there exist constants N_{j-1} and c_{j-1} such that

$$P_k(n) \ge c_{j-1} \frac{k^n}{n \log_k^{\circ j-1}(n) \prod_{i=1}^{j-1} \log_k^{\circ i}(n)}$$

for all $n > N_{j-1}$. By Lemma 6 we have

$$P_k(n) \ge P_k(n,t) \ge \sum_{\substack{|u|=t \\ u \text{ privileged}}} B_k(n,u) \ge \sum_{\substack{|u|=t \\ u \text{ privileged}}} d\frac{k^n}{n^2} = dP_k(t)\frac{k^n}{n^2}.$$

for $n > N_0$. Since $t \le \log_k(n) + 1$, we have that $\frac{1}{\log_k^{\circ i}(t)} \ge \frac{1}{\log_k^{\circ i}(\log_k(n)+1)}$ for all $i \ge 0$. Thus continuing from above we have

$$\begin{aligned}
P_k(n) &\ge dc_{j-1} \frac{k^t}{t \log_k^{\circ j-1}(t) \prod_{i=1}^{j-1} \log_k^{\circ i}(t)} \frac{k^n}{n^2} \ge d_7 \frac{k^{\log_k(n-t)+\log_k(k-1)-\log_k(\ln k)}}{t \log_k^{\circ j-1}(t) \prod_{i=1}^{j-1} \log_k^{\circ i}(t)} \frac{k^n}{n^2} \\
&\ge d_8 \frac{1}{t \log_k^{\circ j-1}(t) \prod_{i=1}^{j-1} \log_k^{\circ i}(t)} \frac{k^n}{n} \\
&\ge d_9 \frac{1}{(\log_k(n)+1) \log_k^{\circ j-1}(\log_k(n)+1) \prod_{i=1}^{j-1} \log_k^{\circ i}(\log_k(n)+1)} \frac{k^n}{n} \\
&\ge c_j \frac{k^n}{n \log_k^{\circ j}(n) \prod_{i=1}^{j} \log_k^{\circ i}(n)}
\end{aligned}$$

for all $n > N_j$ where $N_j > \max(N_0, N_{j-1})$. □

4.2 Upper bound

In Theorem 2 (b) we proved that $C_k(n) \in O(\frac{k^n}{n})$. Since every privileged word is also a closed word, this is also shows that $P_k(n) \in O(\frac{k^n}{n})$. This bound improves on the existing bound on privileged words but it does not show that $P_k(n)$ and $C_k(n)$ behave differently asymptotically. We show that $P_k(n)$ is much smaller than $C_k(n)$ asymptotically by proving upper bounds on $P_k(n)$ that show $P_k(n) \in o(\frac{k^n}{n})$.

Lemma 16. *Let n, t, and k be integers such that $n \geq 2t \geq 2$ and $k \geq 2$. Then*

$$P_k(n,t) \leq P_k(t)A_k(n-2t, 0^t).$$

Proof. The number of length-n privileged words closed by a length-t privileged word is equal to the sum, over all length-t privileged words u, of the number $B_k(n,u)$ of length-n words closed by u. Thus we have that

$$P_k(n,t) = \sum_{\substack{|u|=t \\ u \text{ privileged}}} B_k(n,u).$$

By Lemma 7 we have that $B_k(n,v) \leq A_k(n-2t, 0^t)$ for all length-t words v. Therefore

$$P_k(n,t) = \sum_{\substack{|u|=t \\ u \text{ privileged}}} B_k(n,u) \leq \sum_{\substack{|u|=t \\ u \text{ privileged}}} A_k(n-2t,0^t) \leq P_k(t)A_k(n-2t,0^t).$$

□

Proof of Theorem 3 (b). For $n \geq 2t$ we can use Lemma 16 to bound $P_k(n,t)$. But for $n < 2t$, we can use Corollary 9 and the fact that $P_k(n,t) \leq C_k(n,t)$. We get

$$P_k(n) = \sum_{t=1}^{n-1} P_k(n,t) \leq \sum_{t=1}^{\lfloor n/2 \rfloor} P_k(t)A_k(n-2t,0^t) + nk^{\lceil n/2 \rceil}.$$

The proof is by induction on j. The base case, when $j = 0$, is taken care of by Theorem 2 (b).

Suppose $j > 0$. Then there exist constants N'_{j-1} and c'_{j-1} such that

$$P_k(n) \leq c'_{j-1} \frac{k^n}{n \prod_{i=1}^{j-1} \log_k^{\circ i}(n)}$$

for all $n > N'_{j-1}$. We now bound $P_k(n)$. First we let $N'_j > N'_{j-1}$ be a constant such that the following inequalities hold for all $n > N'_j$. We have

$$
\begin{aligned}
P_k(n) &\le \sum_{t=1}^{\lfloor n/2 \rfloor} P_k(t) A_k(n-2t, 0^t) + nk^{\lceil n/2 \rceil} \\
&\le \sum_{t=N'_j+1}^{\lfloor n/2 \rfloor} c'_{j-1} \frac{k^t}{t \prod_{i=1}^{j-1} \log_k^{\circ i}(t)} \beta_k(t)^{n-2t} + \sum_{t=1}^{N'_j} P_k(t) A_k(n-2t, 0^t) + nk^{\lceil n/2 \rceil} \\
&\le \sum_{t=N'_j+1}^{\lfloor n/2 \rfloor} c'_{j-1} \frac{k^t}{t \prod_{i=1}^{j-1} \log_k^{\circ i}(t)} \left(k - \frac{k-1}{k^{t+1}}\right)^{n-2t} + d_{10} \sum_{t=2}^{N'_j} \left(k - \frac{k-1}{k^{t+1}}\right)^{n-2t} + d_{11} \frac{k^n}{n^2} \\
&\le c'_{j-1} k^n \sum_{t=N'_j+1}^{\lfloor n/2 \rfloor} \frac{1}{k^t t \prod_{i=1}^{j-1} \log_k^{\circ i}(t)} \left(1 - \frac{k-1}{k^{t+2}}\right)^{n-2t} + d_{12} \frac{k^n}{n^2} \\
&\le c'_{j-1} k^n \left(d_{13} \sum_{t=N'_j+1}^{\lfloor \log_k(n) \rfloor} \frac{1}{k^t t \prod_{i=1}^{j-1} \log_k^{\circ i}(t)} \left(1 - \frac{k-1}{k^{t+2}}\right)^{n/2} + \sum_{t=\lfloor \log_k(n) \rfloor + 1}^{\lfloor n/2 \rfloor} \frac{1}{k^t t \prod_{i=1}^{j-1} \log_k^{\circ i}(t)} \right) + d_{12} \frac{k^n}{n^2} \\
&\le c'_{j-1} k^n \left(d_{13} \sum_{t=N'_j+1}^{\lfloor \log_k(n) \rfloor} \frac{1}{k^t t \prod_{i=1}^{j-1} \log_k^{\circ i}(t)} \exp\left(\frac{n}{2} \ln\left(1 - \frac{k-1}{k^{t+2}}\right)\right) \right. \\
&\qquad \left. + \sum_{t=\lfloor \log_k(n) \rfloor + 1}^{\infty} \frac{1}{k^t t \prod_{i=1}^{j-1} \log_k^{\circ i}(t)} \right) + d_{12} \frac{k^n}{n^2}. \qquad (4)
\end{aligned}
$$

The sum on line (4) is clearly convergent. We have

$$
\begin{aligned}
\sum_{t=\lfloor \log_k(n) \rfloor + 1}^{\infty} \frac{1}{k^t t \prod_{i=1}^{j-1} \log_k^{\circ i}(t)} &\le \frac{1}{(\lfloor \log_k(n) \rfloor + 1) \prod_{i=1}^{j-1} \log_k^{\circ i}(\lfloor \log_k(n) \rfloor + 1)} \sum_{t=\lfloor \log_k(n) \rfloor + 1}^{\infty} \frac{1}{k^t} \\
&\le d_{14} \frac{1}{\log_k(n) \prod_{i=1}^{j-1} \log_k^{\circ i}(\log_k(n))} \frac{1}{n} \le d_{14} \frac{1}{n \prod_{i=1}^{j} \log_k^{\circ i}(n)}.
\end{aligned}
$$

Now we upper bound the sum

$$
D_n = \sum_{t=N'_j+1}^{\lfloor \log_k(n) \rfloor} \frac{1}{k^t t \prod_{i=1}^{j-1} \log_k^{\circ i}(t)} \exp\left(\frac{n}{2} \ln\left(1 - \frac{k-1}{k^{t+2}}\right)\right).
$$

It is well-known that $\ln(1-x) \le -x$ for $|x| < 1$. Thus, letting $\alpha = \frac{k-1}{2k^2}$, we have

$$
\exp\left(\frac{n}{2} \ln\left(1 - \frac{k-1}{k^{t+2}}\right)\right) \le \exp\left(-\alpha \frac{n}{k^t}\right).
$$

We reverse the order of the series, by letting $t = \lfloor \log_k(n) \rfloor - t + N'_j + 1$. We also shift the index of the series down by $N'_j + 1$. We have

$$D_n = \sum_{t=0}^{\lfloor \log_k(n) \rfloor - N'_j - 1} \frac{1}{k^{\lfloor \log_k(n) \rfloor - t}(\lfloor \log_k(n) \rfloor - t) \prod_{i=1}^{j-1} \log_k^{\circ i}(\lfloor \log_k(n) \rfloor - t)} \exp\Big(-\alpha \frac{n}{k^{\lfloor \log_k(n) \rfloor - t}}\Big)$$
$$\leq d_{15} \sum_{t=0}^{\lfloor \log_k(n) \rfloor - N'_j - 1} \frac{k^t}{n(\log_k(n) - t) \prod_{i=1}^{j-1} \log_k^{\circ i}(\log_k(n) - t)} \exp(-\alpha k^t)$$
$$\leq d_{15} \frac{1}{n \prod_{i=1}^{j} \log_k^{\circ i}(n)} \sum_{t=0}^{\lfloor \log_k(n) \rfloor - N'_j - 1} \frac{k^t}{\prod_{i=0}^{j-1} \frac{\log_k^{\circ i}(\log_k(n) - t)}{\log_k^{\circ i+1}(n)}} \exp(-\alpha k^t). \tag{5}$$

Suppose β is a positive constant strictly between 0 and 1 such that $\beta \log_k(n)$ is an integer and $\beta \log_k(n) < \lfloor \log_k(n) \rfloor - N'_j - 1$. If $t \leq \beta \log_k(n)$, then $\frac{\log_k^{\circ i}(\log_k(n)-t)}{\log_k^{\circ i+1}(n)} \geq \frac{\log_k^{\circ i+1}(n^{1-\beta})}{\log_k^{\circ i+1}(n)} \geq d'_i$ for some $d'_i > 0$ by Lemma 5. If $t > \beta \log_k(n)$, then $\frac{\log_k^{\circ i}(\log_k(n)-t)}{\log_k^{\circ i+1}(n)} \geq \frac{\log_k^{\circ i}(N'_j+1)}{\log_k^{\circ i+1}(n)}$. We split up the sum in D_n in two parts. One sum with $t \leq \beta \log_k(n)$ and one with $t > \beta \log_k(n)$. Continuing from (5) we get

$$\leq d_{15} \frac{1}{n \prod_{i=1}^{j} \log_k^{\circ i}(n)} \Bigg(\sum_{t=1}^{\beta \log_k(n)} \frac{k^t}{\prod_{i=0}^{j-1} d'_i} \exp(-\alpha k^t) + \prod_{i=0}^{j-1} \bigg(\frac{\log_k^{\circ i+1}(n)}{\log_k^{\circ i}(N'_j + 1)} \bigg) \sum_{t=\beta \log_k(n)+1}^{\lfloor \log_k(n) \rfloor - N'_j - 1} k^t \exp(-\alpha k^t) \Bigg)$$
$$\leq d_{15} \frac{1}{n \prod_{i=1}^{j} \log_k^{\circ i}(n)} \bigg(d_{16} \sum_{t=1}^{\infty} t \exp(-\alpha t) + d_{17} \prod_{i=1}^{j} \log_k^{\circ i}(n) \sum_{t=kn^\beta}^{\infty} t \exp(-\alpha t) \bigg).$$

The first and second sum are both clearly convergent. It is also easy to show that both of them can be bounded by a constant multiplied by the first term. Thus, we have that

$$D_n \leq d_{15} \frac{1}{n \prod_{i=1}^{j} \log_k^{\circ i}(n)} \bigg(d_{18} + d_{19} \prod_{i=1}^{j} \log_k^{\circ i}(n) \frac{kn^\beta}{\exp(\alpha k n^\beta)} \bigg) \leq d_{20} \frac{1}{n \prod_{i=1}^{j} \log_k^{\circ i}(n)}.$$

Putting everything together and continuing from line (4), we get

$$P_k(n) \leq c' k^n \bigg(d_{13} D_n + d_{14} \frac{1}{n \prod_{i=1}^{j} \log_k^{\circ i}(n)} \bigg) + d_{12} \frac{k^n}{n^2} \leq c'_j \frac{k^n}{n \prod_{i=1}^{j} \log_k^{\circ i}(n)}$$

for some constant $c'_j > 0$. □

5 Open problems

We conclude by posing some open problems.

1. In this paper we showed that $C_k(n) \in \Theta(\frac{k^n}{n})$. In other words, we showed that $C_k(n)$ can be bounded above and below by a constant times k^n/n for n sufficiently large. Can we do better than this? Does the limit

$$\lim_{n\to\infty} \frac{C_k(n)}{k^n/n}$$

exist? If it does exist, what does the limit evaluate to? Can one find good bounds on the limit?

2. In this paper we also gave a family of upper and lower bounds for $P_k(n)$. But for every $j \geq 0$, the upper and lower bounds on $P_k(n)$ are asymptotically separated by a factor of $1/\log_k^{\circ j}(n)$. Does there exist a $g(n)$ such that $P_k(n) \in \Theta(\frac{k^n}{g(n)})$? If such a function $g(n)$ exists, then does the limit

$$\lim_{n\to\infty} \frac{P_k(n)}{k^n/g(n)}$$

exist?

6 Acknowledgements

Thanks to Jeffrey Shallit for introducing me to this problem and for helpful discussions and suggestions.

References

[1] G. Badkobeh, G. Fici, and Z. Lipták. On the number of closed factors in a word. In Adrian-Horia Dediu, Enrico Formenti, Carlos Martín-Vide, and Bianca Truthe, editors, *Language and Automata Theory and Applications*, pp. 381–390, Cham, 2015. Springer International Publishing.

[2] M. Bucci, A. De Luca, and G. Fici. Enumeration and structure of trapezoidal words. *Theoret. Comput. Sci.* **468** (2013), 12–22.

[3] A. Carpi and A. de Luca. Periodic-like words, periodicity, and boxes. *Acta Informatica* **37**(8) (2001), 597–618.

[4] A. De Luca and G. Fici. Open and closed prefixes of Sturmian words. In J. Karhumäki, A. Lepistö, and L. Zamboni, editors, *Combinatorics on Words*, pp. 132–142, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.

[5] G. Fici. A classification of trapezoidal words. In P. Ambrož, Š. Holub, and Z. Masáková, editors, *8th International Conference on Words, WORDS 2011*, Vol. 63 of *Electronic Proceedings in Theoretical Computer Science*, pp. 129–137. 2011.

[6] G. Fici. Open and closed words. *Bull. European Assoc. Theor. Comput. Sci.* , No. 123, (2017), 138–147.

[7] M. Forsyth, A. Jayakumar, J. Peltomäki, and J. Shallit. Remarks on privileged words. *Internat. J. Found. Comp. Sci.* **27**(4) (2016), 431–442.

[8] E. Gilbert. Synchronization of binary messages. *IRE Trans. Info. Theory* **6**(4) (1960), 470–477.

[9] L. J. Guibas and A. M. Odlyzko. Maximal prefix-synchronized codes. *SIAM J. Appl. Math.* **35**(2) (1978), 401–418.

[10] L. J. Guibas and A. M. Odlyzko. Periods in strings. *J. Combin. Theory Ser. A* **30**(1) (1981), 19–42.

[11] L. J. Guibas and A. M. Odlyzko. String overlaps, pattern matching, and nontransitive games. *J. Combin. Theory Ser. A* **30**(2) (1981), 183–208.

[12] M. Jahannia, M. Mohammad-Noori, N. Rampersad, and M. Stipulanti. Closed Ziv–Lempel factorization of the m-bonacci words. *Theoret. Comput. Sci.* **918** (2022), 32–47.

[13] J. Kellendonk, D. Lenz, and J. Savinien. A characterization of subshifts with bounded powers. *Discrete Math.* **313**(24) (2013), 2881–2894.

[14] J. Nicholson and N. Rampersad. Improved estimates for the number of privileged words. *J. Integer Sequences* **21** (2018), Article 18.3.8 (electronic).

[15] N. J. A. Sloane et al. OEIS Foundation Inc. (2022), The On-Line Encyclopedia of Integer Sequences, `https://oeis.org`.

[16] J. Peltomäki. Introducing privileged words: Privileged complexity of Sturmian words. *Theoret. Comput. Sci.* **500** (2013), 57–67.

[17] J. Peltomäki. Privileged factors in the Thue–Morse word—a comparison of privileged words and palindromes. *Disc. Appl. Math.* **193** (2015), 187–199.

[18] J. Rukavicka. Upper bound for the number of closed and privileged words. *Inform. Process. Lett.* **156** (2020), 105917.

[19] J. Rukavicka. Upper bound for the number of privileged words. *Discrete Math.* **346**(1) (2023), 113164.

[20] L. Schaeffer and J. Shallit. Closed, palindromic, rich, privileged, trapezoidal, and balanced words in automatic sequences. *Electronic J. Combinatorics* **23**(1) (2016), P1.25 (electronic).