Enumeration of certain words *

By K. H. KIM and F. W. ROUSH

1. Introduction

Both the content and methods of this paper are closely related to those of [2]. Let F denote a free semigroup on generators $x_1, x_2, \ldots, x_m$. We wish to enumerate, for fixed n and k, the number of length n words which do not have any segment which is the square of a length k word.

One reason for considering this problem is that it is related to the more difficult problem of enumerating words not containing any segment which is a square word. GOULDEN and JACKSON [1, Corollary 4.1] have obtained our Theorem 1 by completely different methods.

2. A recursion formula

Definition 1. Let F^1 be a *free monoid.* Let $w_1, w_2 \in F^1$.
(1) $w_1 | w_2$ if $xw_1y = w_2$ for some $x, y \in F^1$.
(2) $w_1 |_i w_2$ if $w_2 = w_1 y$ for some $y \in F^1$.
(3) $w_1 |_f w_2$ if $w_2 = xw_1$ for some $x \in F^1$.
(4) $|w_1|$ denotes the length of w_1; $|1| = 0$.
(5) If $w_1 \neq 1$, $\hat{w}_1$ is the unique word such that

$$\hat{w}_1 |_i w_1, \; |\hat{w}_1| = |w_1| - 1.$$

Definition 2. Let F be a *free semigroup* on $m > 1$ generators and F^1 the associated free monoid. Fix $k > 0$.

To any word w in F^1 we assign a length $k-1$ (0, 1)-vector v as follows. The number v_j is 1 if and only if the $n-k+j+1$ letter of w equals the $n-2k+j+1$ letter of w, and $n-2k+j+1 > 0$. We assign an integer $a(w)$ to w by stating that $a(w)$ is the length of the longest terminal sequence of 1 components of v. Let $S(n, c)$ for $0 \leqq c \leqq k-1$ denote the set of length n words w_1 of F^1 such that $a(w_1) = c$ and if $|w_2| = k$ it is false that $(w_2)^2 | w_1$.

* This work was supported by Alabama State University Faculty Research Grant R—78—6.

In other terms, given a word w_1 consider the terminal segment w_3 of length $2k-1$ of w_1. Then $a(w_1)$ is the length of the longest initial segment of w_3 which equals a final segment of w_3, or is $k-1$ if this length exceeds $k-1$. The set $S(n, c)$ is the set of words w of length n such that $a(w)=c$ and w is not divisible by the square of a word of length k.

Theorem 1. *For* $n>k, 0<c<k,$

$$|S(n, c)| = |S(n-1, c-1)|,$$

$$|S(n, 0)| = \sum_{j=0}^{k-1} (m-1)|S(n-1, j)|.$$

Furthermore, $|S(n, 0)|=m^n$, $|S(n, c)|=0$ *for* $n\leqq k$, $0<c<k$.

Proof. Suppose $x\in S(n, c)$. Then $\hat{x}\in S(n, c-1)$. (Note that if $c=k-1$, the $n-k$ and $n-2k$ letters of x must differ, else x would be divisible by the square of a length k word.) And if $y\in S(n, c-1)$ there is one and only one $x\in S(n, c)$ such that $\hat{x}=y$. Namely the last letter of x must equal the $n-k$ letter of y. Thus $|S(n, c)|=|S(n, c-1)|$. Also the function $x\to\hat{x}$ maps $S(n, 0)$ into $\bigcup_{j=0}^{k-1} S(n-1, j)$. Each element y of the latter set arises from exactly $(m-1)$ elements of $S(n, 0)$. Namely we can add to y any letter except the $(n-k)^{\text{th}}$. Therefore

$$|S(n, 0)| = (m-1) \sum_{j=0}^{k-1} |S(n-1, j)|.$$

This proves the theorem.

This formula can be recast into a matrix form. Let M be the $k\times k$ matrix

$$\begin{bmatrix} m-1 & 1 & 0\dots 0 \\ m-1 & 0 & 1\dots 0 \\ m-1 & 0 & 0\dots 0 \\ & \dots & \\ m-1 & 0 & 0\dots 0 \end{bmatrix}.$$

Let u be the vector $(m^k, 0, 0, \dots, 0)$. Then $|S(n, c)|$ is the $c+1$ component of uM^{n-k}, for $n\geqq k$. The characteristic polynomial of M is

$$P(z) = z^k-(m-1)(z^{k-1}+z^{k-2}+\dots+z+1).$$

Definition 3. For $n<k$ put $f(n)=0$. For $n\geqq k$, let $f(n)$ be

$$\sum_{j=0}^{k-1} |S(n, j)|.$$

Theorem 2. *The generating function of* $f(n)$ *is*

$$\frac{m^k z^k(1-z^k)}{1-mz+(m-1)z^{k+1}}.$$

Proof. As in [2], the generating function of $f(n)$ must have the form

$$\frac{q(z)z^k}{z^k p\left(\frac{1}{z}\right)}$$

where $q(z)$ is some polynomial of degree at most $k-1$. In degree less than $2k$, this quotient must be

$$z^k m^k(1+mz+\ldots+m^{k-1}z^{k-1}).$$

Therefore $q(z)$ is the portion of

$$m^k(1-(m-1)z-\ldots-(m-1)z^k)(1+mz+\ldots+m^{k-1}z^{k-1})$$

having degrees no more than k. A computation gives the formula above (note that $(1-z)$ can be cancelled from numerator and denominator).

3. Asymptotic values of $f(n)$

Lemma 3. *The equation $P(z)=0$ has a unique positive real root r_k of multiplicity 1. This root exceeds the absolute value of any other root. We have $r_k > r_{k-1}$ and*

$$\lim_{k\to\infty} r_k = m.$$

Moreover

$$m-\frac{m-1}{m^k} > r_k > m-\frac{m-1}{r_j^k}$$

for $j<k$.

Proof. If $P(z)=0$ then $u=\frac{1}{z}$ satisfies

$$u^{k+1} = \frac{mu-1}{m-1}$$

as does $u=1$. However no straight line can cut the curve $y=x^{k+1}$, $x>0$ in more than two places since $y=x^{k+1}$ is concave upwards. Therefore 1 and $\frac{1}{r_k}$ are the only positive solutions of

$$u^{k+1} = \frac{mu-1}{m-1}.$$

Therefore $P(z)=0$ has only one positive real solution r_k and $r_k>1$. Differentiation shows that r_k has multiplicity 1. Let z be a root of $P(z)=0$ which is negative or complex. Then

$$z^k = (m-1)(z^{k-1}+z^{k-2}+\ldots+1)$$

implies

$$|z|^k < (m-1)(|z|^{k-1}+|z|^{k-2}+\ldots+1) .$$

So $|z|<r_k$ because for $|z|\geqq r_k$ we have $P(|z|)>0$. We have

$$r_k^{k-1} = (m-1)\left(r_k^{k-2}+\ldots+1+\frac{1}{r_k}\right) > (r_k^{k-2}+\ldots+1)(m-1)$$

and

$$r_{k-1}^{k-1} = (m-1)(r_{k-1}^{k-2}+\ldots+1).$$

This implies $r_k>r_{k-1}$. It also implies that

$$r = \lim_{k\to\infty} r_k$$

satisfies

$$r = (m-1)\frac{1}{1-\frac{1}{r}}.$$

Therefore $r=m$.

From

$$z^k = (m-1)\frac{z^k-1}{z-1}$$

at $z=r_k$ we have

$$z = 1+(m-1)\left(1-\frac{1}{z^k}\right).$$

This implies the last inequality of the lemma. This proves the lemma.

Theorem 4. *The asymptotic value of $f(n)$ is*

$$\frac{m^k(1-u^k)u^{k-n-1}}{m-(m-1)(k+1)u^k}$$

where $u=\dfrac{1}{r_k}$.

Proof. Expand the generating function in partial fractions. All other terms will be insignificant compared with the term

$$\frac{A}{1-r_k z}.$$

This term can computed by letting z approach $\dfrac{1}{r_k}$ in the generating function. This proves the theorem.

Abstract

We study the number of words of length n, in m generators, divisible by the square of a length k word. We find a recursion formula, the generating function, and the asymptotic value of this number.

MATHEMATICS RESEARCH GROUP
ALABAMA STATE UNIVERSITY
MONTGOMERY, ALABAMA 36101
U.S.A.

Reference

[1] GOULDEN, I. M. and D. M. JACKSON, An inversion theorem for cluster decompositions of sequences with distinguished subsequences, Univ. of Waterloo, Dept. of Combinatorics and Optimization Research Report CORR 78/24, August 1978.
[2] KIM, K. H., M. S. PUTCHA and F. W. ROUSH, Some combinatorial properties of free semigroups, *J. London Math. Soc.* (2), v. 16, 1977, pp. 397—402.

(Received Nov. 9, 1978)