

THE ELEMENTS OF DISSERTATION

A Thesis
Submitted to
the Temple University Graduate Board

in Partial Fulfillment
of the Requirements for the Degree of
MASTER OF SCIENCE

by
Perry H. Disdainful
January, 2000

Examining Committee Members:

Ignatius Arrogant, Mathematics, Advisory Committee Chair
Severus Snobb, Mathematics, Temple University
Sophus Knowitall, Mathematics, Temple University
Cerberus Angerbight, Biology, Hades University

©

by

Perry H. Disdainful

January, 2000

All Rights Reserved

ABSTRACT

THE ELEMENTS OF DISSERTATION

Perry H. Disdainful

MASTER OF SCIENCE

Temple University, January, 2000

Ignatius Arrogant, Mathematics, Chair

Theses have elements. Isn't that nice?

ACKNOWLEDGEMENTS

I want to “thank” my committee, without whose ridiculous demands, I would have graduated so, so, very much faster.

To myself,
Perry H. Disdainful,
the only person worthy of my company.

TABLE OF CONTENTS

ABSTRACT	iv
ACKNOWLEDGEMENT	v
DEDICATION	vi
LIST OF FIGURES	viii
LIST OF TABLES	ix
1 SEQUENCES OF WORDS	1
1.1 Words	1
1.2 Avoiding the Bad	2
1.3 Substantive Sequences	3
1.4 The Naive Approach	4
1.5 The Goulden-Jackson Method	5
2 BINARY CUBE-FREE WORDS	6
2.1 Introduction	6
2.2 Applying the Goulden-Jackson Method	6
2.2.1 The Sequence of Binary Cube-Free Words of length up to 47	7
2.2.2 The ‘Connective Constant’	7
2.3 Lower-Bounds and the Brinkhuis Method	8
2.3.1 Lower Bounds for Square-free Ternary Words	8
2.3.2 Lower Bounds for Cube-Free Binary Words	9
3 TABLE STUFF	11
4 FIGURE PLACEMENT	13
REFERENCES	15

LIST OF FIGURES

4.1	A first figure.	13
4.2	A second figure.	14

LIST OF TABLES

3.1	A small table.	11
3.2	A typical table.	12

CHAPTER 1

SEQUENCES OF WORDS

1.1 Words

Words surround us. Not just in the literal sense of the words on billboards, road signs, cereal packets, and in books and magazines, but also in a more abstract sense. Our DNA is defined by a word over the language of nucleotides. The bar codes on our groceries are words in the computer language of zeroes and ones. Further, in mathematics there are words that avoid certain patterns, such as repeating blocks which can be explored purely for their own interest, and some that have applications in such areas as the study of linear polymer molecules in chemical physics.

In order to explore the behavior of such a wide range of words we must first introduce a format by which words are defined, and some basic terminology that will be used throughout this work. My choice of notation is based on my frequent reliance on Maple to perform calculations.

Notation 1.1 *Let V be the alphabet over which our language is defined.*

E.g. in English $V = \{a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, q, r, s, t, u, v, w, x, y, z\}$.
In computing $V = \{0, 1\}$.

Definition 1.1 *A **word**, w , over the alphabet V is an ordered sequence of letters from V , $w = [w_1, w_2, \dots, w_n]$ where $w_i \in V$ for $1 \leq i \leq n$.*

E.g. the English word “alphabet” becomes $[a, l, p, h, a, b, e, t]$.

Notation 1.2 V^* is the set of all possible words over the alphabet V .

Definition 1.2 A **sub-word** of w is any of the $\binom{n+1}{2}$ possible sub-sequences $[w_i, w_{i+1}, \dots, w_j]$ where $1 \leq i \leq j \leq n$.

Thus $[a, l, p]$, $[h, a]$ and $[b, e, t]$ are all sub-words of $[a, l, p, h, a, b, e, t]$.

Notation 1.3 The **empty word** is considered to be a sub-word of all words and belongs to V^* for every V . It will be denoted $[]$

Definition 1.3 The **length** of a word $l(w)$ is the number of letters in the word, counting multiplicity.

E.g. $l([a, l, p, h, a, b, e, t]) = 8$. Note $l([]) = 0$.

One of the main areas of research into words is their limiting behavior. That is if a_n is the number of words in our language of length n we want to find $\mu := \lim_{n \rightarrow \infty} [a_n]^{1/n}$, if it exists.

Clearly if no constraints are put upon our choice of words and if k is the number of letters in our alphabet V then $a(n) = k^n$ and hence $\mu = k$. This leads us to believe our quest for limits will not prove fruitless.

Often it is useful to use the model $a_n = n^\theta \mu^n$. Zinn’s method can be used to obtain good approximations of this type.

1.2 Avoiding the Bad

Most of the sequences a_n considered in this text are ones whose words avoid specific sub-words. We consider the sub-words we wish to avoid as the bad words (or mistakes), and the set of all such words will be denoted B . The set of all bad words up to length k will be denoted B_k .

As an example consider the case of binary square-free words. That is words over a two letter alphabet that avoid any non-trivial sub-word being repeated directly after itself. In this case $B_4 = \{[0, 0], [1, 1], [0, 1, 0, 1], [1, 0, 1, 0]\}$.

It should be noted that B and B_k are always minimal in the sense that no member of B (or B_k) is a sub-word of any other member of B (or B_k). In the above example note $[1, 1, 1, 1]$ is omitted from B_4 because it contains $[1, 1]$ as a sub-word.

In fact in this case $B = B_4$ and $a_n = [1, 2, 2, 2, 0, 0, 0, \dots]$, which is not a very interesting sequence. The more interesting case of ternary square-free words is discussed by Noonan [6].

1.3 Substantive Sequences

Many of the sequences we will be discussing are sub-multiplicative. That is that $a_{n+m} \leq a_n a_m$. In sequences where $a_n \neq 0$ we have that $\log(a_{n+m}) \leq \log(a_n) + \log(a_m)$ which shows that the sequence $\{\log(a_n)\}$ is subadditive ($c_{n+m} \leq c_n + c_m$). This fact can be used to show that the μ exists and is in fact the $\inf a_n^{(1/n)}$

Lemma 1.1 *Let $\{c_n\}$ be a subadditive sequence of real numbers. Then the $\lim_{n \rightarrow \infty} \frac{c_n}{n}$ exists and equals $\inf_{n \geq 1} \frac{c_n}{n}$.*

The above lemma is attributed to Hammersley and Morton (1954).

Proof of Lemma: Let $C_k = \max_{1 \leq r \leq k} c_r$. Then for any given n we can find j such that $n = jk + r$ with $1 \leq r \leq k$.

Using the subadditivity of $\{c_n\}$ we obtain

$$c_n \leq jc_k + c_r \leq \frac{n}{k}c_k + C_k \quad (1.1)$$

Then we divide both sides by n and take the $\limsup_{n \rightarrow \infty}$ to obtain

$$\limsup_{n \rightarrow \infty} \frac{c_n}{n} \leq \limsup_{n \rightarrow \infty} \left(\frac{c_k}{k} + \frac{C_k}{n} \right) \leq \frac{c_k}{k} \quad (1.2)$$

Finally we take the $\liminf_{k \rightarrow \infty}$ and obtain that the $\limsup \leq \liminf$ thus proving the limit exists.

As the limit exists it equals the \limsup and so as this is less than $\frac{a_k}{k}$ for all k we obtain

$$\lim_{n \rightarrow \infty} \frac{c_n}{n} = \inf_{n \geq 1} \frac{c_n}{n} \quad (1.3)$$

Theorem 1.1 *If $\{a_n\}$ is a sequence of positive terms for which $a_{n+m} \leq a_n a_m$ then $\mu = \lim_{n \rightarrow \infty} a_n^{\frac{1}{n}}$ exists. Further $\mu \leq a_n^{\frac{1}{n}}$.*

Proof: As discussed above $a_{n+m} \leq a_n a_m$ implies that the sequence $\{\log a_n\}$ is subadditive. This means $\log \mu = \lim_{n \rightarrow \infty} \frac{\log a_n}{n} = \lim_{n \rightarrow \infty} \log a_n^{\frac{1}{n}}$ exists and further $\log \mu = \inf_{n \geq 1} \frac{\log a_n}{n} = \inf_{n \geq 1} \log a_n^{\frac{1}{n}} \leq \log a_n^{\frac{1}{n}}$ for all n . And this gives the required results.

1.4 The Naive Approach

At this point we are only considering linear sequences. Later in this Thesis we will investigate the case of cyclic sequences.

For any given word we define its k -weight at follows:

$$W_k([w_1, w_2, \dots, w_n]) = s^n \prod_{i=1}^k \prod_{j=1}^{n-k+1} x[w_j \dots w_{j+k-1}] \quad (1.4)$$

For example the 3-weight of the word apple would be

$$s^5 x[a]x[p]^2 x[l]x[e]x[a, p]x[p, p]x[p, l]x[l, e]x[a, p, p]x[p, p, l]x[p, l, e]$$

So that the coefficient of s^n will give us the number of words of length n and when necessary their form. This means are goal becomes to find the generating function that has all words of length n (or often just the number of them) that meet our criteria as the coefficient of s^n .

One method for doing this is to use a matrix, A , to analyze the interaction between all possible blocks of length k then by taking $(1 - A)^{-1}$ and adding all the resulting entries we obtain a generating function for all words over the

chosen alphabet. We then set any blocks that are disallowed equal to zero and obtain the generating function for the desired set of words.

This method I call the Naive Approach because it produces all possible words without taking into account the bad words until the very end. For example if we were to take the English alphabet and look for all words that did not contain any bad “4-letter” words we would need a matrix that was 26^4 by 26^4 and worse yet need to find the inverse of such a matrix a very slow task, even for a computer. Thus this approach is only useful in very small cases and as a check for our clever techniques, like the Goulden-Jackson Method.

1.5 The Goulden-Jackson Method

One method used throughout this dissertation is the Goulden-Jackson Cluster Method [4]. This method can be used to find the generating function $f(s) = \sum_{n=0}^{\infty} a_n s^n$. In many cases we can not find $f(s)$ explicitly as we are looking at infinite sets of mistakes, but we can obtain $f_k(s)$ which gives correct values for a_n when $n \leq k$ and good over estimates for $n > k$.

We will discuss briefly this method, for a more in depth explanation see [4].

CHAPTER 2

BINARY CUBE-FREE WORDS

2.1 Introduction

Definition 2.1 *A word is **cube-free** if it contains no factors of the form xxx , where x is any non-empty word.*

E.g. The cube-free words of length 3 over the alphabet $\{a, b\}$ are

$\{[a, a, b], [a, b, a], [a, b, b], [b, a, a], [b, a, b], [b, b, a]\}$

My Maple package Cubefree (available from <http://www.math.temple.edu/~anne/cubefree.html>) can be used to derive cube-free words on any given alphabet up to the required length. The number of binary cube-free words of length at most n for $0 \leq n \leq 47$ are given below.

2.2 Applying the Goulden-Jackson Method

These results were obtained by applying the Goulden-Jackson Method with all cubes of length at most 45 as the input mistakes.

2.2.1 The Sequence of Binary Cube-Free Words of length up to 47

1, 2, 4, 6, 10, 16, 24, 36, 56, 80, 118, 174, 254, 378, 554, 802, 1168, 1716, 2502, 3650, 5324, 7754, 11320, 16502, 24054, 35058, 51144, 74540, 108664, 158372, 230800, 336480, 490458, 714856, 1041910, 1518840, 2213868, 3226896, 4703372, 6855388, 9992596, 14565048, 21229606, 30943516, 45102942, 65741224, 95822908, 139669094.

2.2.2 The ‘Connective Constant’

Let a_n be the number of cube-free words of length n . Brandenburg [1] proved that for $n > 18$

Lemma 2.1 $\{a_n\}$ is sub-multiplicative.

Proof: Given a cube free word of length $n + m$ if we split it into the first n letters and the last m letters both of these words must be cube-free or the original word was not. Hence $a_{n+m} \leq a_n a_m$.

It is also worth noting that when we adjoin two cube-free words we do not necessarily obtain a cube-free word so this is not multiplicative.

Theorem 2.1 $\mu = \lim_{n \rightarrow \infty} a_n^{1/n}$ exists and $\mu = \liminf_{n \rightarrow \infty} a_n$.

Proof: See 1.1.

$$2 \times 1.080^n < 2 \times 2^{\frac{n}{9}} \leq a_n \leq 2 \times 1251^{\frac{n-1}{17}} < 1.315 \times 1.522^n \quad (2.1)$$

Thus $1.080 \leq \mu \leq 1.522$

Using the ‘memory-45’ analog (i.e. the corresponding sequence that enumerates words that avoid cubes x^3 , with $\text{length}(x) \leq 15$), that was generated using the Maple package, up to word-length 300, we find the rigorous upper bound $\mu < 1.457579200596766$, which improves on Brandenburg’s result.

Using Zinn's method, we also found that, assuming that $a_n \sim n^\theta \mu^n$, then $\mu \approx 1.457$, and $\theta \approx 0$. Hence it is reasonable to conjecture that $a_n \sim \mu^n$, where $\mu := \lim_{n \rightarrow \infty} a_n^{1/n} \approx 1.457$.

2.3 Lower-Bounds and the Brinkhuis Method

2.3.1 Lower Bounds for Square-free Ternary Words

Jan Brinkhuis [2] obtained a lower bound for the number of square-free ternary words in the following way. He found a pair of words, $U0, V0$, on $\{0, 1, 2\}$ and from these forms $U1, V1$ and $U2, V2$ all with the following property. If W is a square-free word over $\{0, 1, 2\}$, and $S(W)$ is obtained by replacing all the 0's in W with UO or VO , the 1's with $U1$ or $V1$ and the 2's with $U2$ or $V2$ then $S(W)$ is also square-free.

Lemma 2.2 *If we can find $U0, V0, U1, V1, U2, V2$ that satisfy the above condition and are of length k then $\mu \geq 2^{\frac{1}{k-1}}$.*

Proof: As we have two choices of what to substitute for each of the letters of W

$$a_{kn} \geq 2^n a_n \quad (2.2)$$

Thus

$$a_{kn}^{\frac{1}{kn}} \geq 2^{\frac{1}{k}} (a_n^{\frac{1}{n}})^{\frac{1}{k}} \quad (2.3)$$

and taking the limit with respect to n we obtain

$$\mu \geq 2^{\frac{1}{k}} \mu^{\frac{1}{k}} \quad (2.4)$$

which simplifies to

$$\mu \geq 2^{\frac{1}{k-1}} \quad (2.5)$$

Brinkhuis chose words that were palindromes and obtained $U1$ from $U0$ by adding 1 mod 3 to each letter of $U0$ and $U2$ is obtained from $U0$ by adding 2 mod 3 to each letter of $U0$. Likewise for $V1$ and $V2$. He found (by

hand) such a Brinkhuis pair ($U0$ and $V0$) of length 24. Giving lower bound of $\mu \geq 2^{\frac{1}{23}} = 1.030595545$

Zeilberger and his servant Ekhad [8] removed the palindromic requirement and computerized the search for good pairs. They thus found a Brinkhuis pair of length 18, and so improved the lower bound to $\mu \geq 2^{\frac{1}{17}} = 1.04162$.

In their paper Zeilberger and Ekhad note that the relationship between $U0, U1$ and $U2$ and between $V0, V1$ and $V2$ is not necessary and it is with this comment in mind that I began my adaptation of the Brinkhuis method to cube-free words.

2.3.2 Lower Bounds for Cube-Free Binary Words

Theorem 2.2 *The number of n -letter binary cube-free words is greater than $2^{n/8}$.*

This result can be obtained as a corollary of Brandenburg's result, but as my method is different from his I will give the full details.

The goal is to find binary words $U0, U1, V0, V1$ of minimal length such that if we take a cube-free word W over the alphabet $\{0, 1\}$ and substitute $U0$ or $V0$ for the zeros and $U1$ or $V1$ for the ones the resulting word $S(W)$ will also be cube-free.

Lemma 2.3 *If $U0, V0, U1$, and $V1$ satisfy the following conditions and if W is cube-free then $S(W)$ is cube-free.*

- 1) *All legitimate triples of $U0, V0, U1, V1$ are cube-free*
- 2) *None of $U0, V0, U1, V1$ are non-trivial sub-words of all the possible pairs of $U0, V0, U1, V1$*

Proof:

Clearly as $U0, V0, U1$, and $V1$ meet condition 1 then if W is cube-free and of length at most 3 then $S(W)$ is cube-free.

So if $S(W)$ contains a cube it has length greater than 3. For such a word to contain a cube the pattern of at least one of $U0, V0, U1$, and $V1$ must

be repeated elsewhere in $S(W)$. If every time such a repetition occurs it is as $U0, V0, U1$, and $V1$ respectively then the original word W cannot of been cube-free (contrary to assumptions). So the only way the repeat can occur is as a sub-word of a pair of concatenated words, but condition 2 eliminates this possibility. Therefore $S(W)$ is cube-free whenever W is.

Lemma 2.4 *If we can find $U0, V0, U1, V1$ that satisfy the above condition and are of length k then $\mu = \lim_{n \rightarrow \infty} a_n^{1/n} \geq 2^{\frac{1}{k-1}}$. Where a_n is the number of cube-free words of length n .*

Proof: As for the lemma 2.2 in the square-free case.

Proof of Theorem: It is easily verified (by hand , or more quickly by computer) that $U0 = [0, 1, 1, 0, 0, 1, 1, 0, 1]$, $V0 = [0, 1, 1, 0, 1, 0, 0, 1, 0]$, $U1 = [1, 0, 0, 1, 1, 0, 0, 1, 0]$, and $V1 = [1, 0, 0, 1, 0, 1, 1, 0, 1]$ satisfy the conditions of the lemma. Hence $a(n) \geq 2^{1/8} \approx 1.09$

It should be noted that our words are not palindromic, but $U1$ and $V1$ can be obtained by switching 1's and 0's and vice-versa in $U0$ and $V0$. Removing this condition does not seem to produce any shorter choices for $U0, V0, U1$ and $V1$

CHAPTER 3

TABLE STUFF

Below are two tables to illustrating how tables may be placed in documents.

Table 3.1: A small table.

Title	Author
Vector bundles and you.	Unknown
The Great Gatsby	F. Scott Fitzgerald

Table 3.2: A typical table.

100,000 Samples from $N(0,1)$		
U	C	No. of points at and above U
4.0009	-0.5983	6.0000
3.9402	-1.5813	7.0000
3.8931	-1.6662	8.0000
3.8159	-2.4016	9.0000
3.8081	-1.1640	10.0000
3.7955	-0.7658	11.0000
3.7878	-0.5049	12.0000
3.5760	-0.5738	23.0000
3.5664	-0.5083	24.0000
3.5466	-0.5428	25.0000

CHAPTER 4

FIGURE PLACEMENT

Below are some figure placement examples.

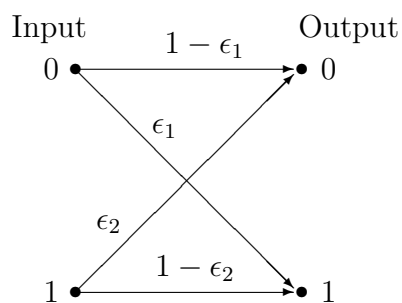


Figure 4.1: A first figure.

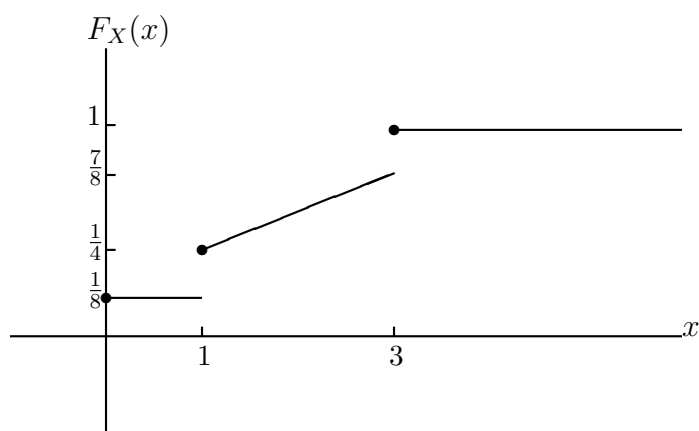


Figure 4.2: A second figure.

REFERENCES

- [1] Brandenburg, F.J. 1983. *Uniformly Growing k -th Power-Free Homomorphisms*. Theoretical Computer Science **23**, 69-82.
- [2] Brinkhuis, J. 1983. *Non-Repetitive Sequences on Three Symbols*. Quart. J. Math. Oxford **2**, no. 34: 145-149.
- [3] Burstein, A., and Wilf, H.S. 1997. *On Cyclic Strings Without Long Constant Blocks*. Fibonacci Quart. **35**, no. 3: 240-247.
- [4] Goulden, I.P., and Jackson, D.M. 1979. *An Inversion Theorem for Cluster Decompositions of sequences with Distinguished Subsequences*. J. London Math. Soc. **2**, no. 20: 567-576.
- [5] Madras, N., and Slade, G. 1996. *The Self-Avoiding Walk*. Birkhauser, Boston.
- [6] Noonan, J., and Zeilberger, D. 1999 *The Goulden-Jackson Cluster Method: Extensions, Applications and Implementations*. J. Difference Eq. Appl. **5** : 355-377.
- [7] Révész, György E. 1983 *Introduction to Formal Languages*. Dover, New York.
- [8] Zeilberger, D., Ekhad, S.B. 1998. *There are more than $2^{n/17}$ n -letter Ternary Square-Free Words*. Journal of Integer Sequences.