2006

On Prime Generation Through Primitive Divisors Of Recurrence Sequences

Richard Russell
University of Central Florida

On Prime Generation Through Primitive Divisors of Recurrence Sequences

by

RICHARD L. RUSSELL
B.S. Mathematics University of Mississippi, 2004

A thesis submitted in partial fulfillment of the requirements
for the degree of Master of Science
in the Department of Mathemtics
in the College of Arts and Sciences
at the University of Central Florida
Orlando, Florida

Spring Term
2006

ABSTRACT

We examine results concerning the generation of primes in certain types of integer sequences. The sequences discussed all have a connection in that each satisfies a recurrence relation. Mathematicians have speculated over many centuries that these sequences contain an infinite number of prime terms, however no proof has been given as such. We examine a less direct method of showing an infinitude of primes in each sequence by showing that the sequences contain an infinite number of terms with primitive divisors.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

1.1: Historical Introduction to Counting Prime Numbers

The prime numbers are the building blocks of the multiplicative structure of the natural numbers, and as such, the theory of prime numbers has fascinated mathematicians for centuries. More than 2000 years ago, in Book IX of his well-known ***Elements***, Euclid proved that any integer can be written as a product of primes and also that there are an infinite number of primes. Although several other proofs of the infinitude of primes have been written through the centuries, Euclid's work holds up today as an excellent model of mathematical reasoning and is still taught in classrooms around the world.

During the 17^{th} and 18^{th} centuries, the research concerning prime numbers turned to the study of primes of certain forms. Some of the work of that period included Fermat's studies of primes of the form $4n+1$, which he proved could always be written as the sum of two squares. He also stated that numbers of the form 2^n+1, which we know today as *Fermat numbers*, are always prime if n is a power of two. However, Euler disproved this statement 100 years later by showing that $2^{32}+1$ is not prime.

Many early problems from the study of primes remain open today, including:

- Are there infinitely many pairs of primes only two integers apart? (*The Twin Prime Conjecture*)

- Goldbach's conjecture that every even integer greater than or equal to 4 can be written as the sum of two primes.
- Are there infinitely many prime Fermat numbers?
- Are there infinitely many primes of the form $2^n - 1$ (*Mersenne primes)*?
- Are there infinitely many primes of the form $n^2 + 1$?

In 1914, the American mathematician D. N. Lehmer published a table listing all of the primes less than 10 million. There are 664,579 primes less than 10 million, or about 6.5% of all integers less than 10 million. D. H. Lehmer, the son of D. N. Lehmer, later calculated the number of primes less than 10 billion and found that there are exactly 455,052,511 such primes, or about 4.5%.[16] A careful examination of any table of primes will show that they are distributed in a very irregular fashion, and this is due to the fact that no simple formula exists for producing all the primes.

There do exist certain expressions which yield a number of primes. For example, $x^2 - x + 41$ gives a prime for $x = 0, 1, 2, ..., 40$. However, no such simple formula can give a prime for all *x*. In 1837, Dirichlet proved that, if *a* and *b* are positive coprime integers, then $ax + b$ gives infinitely many primes as *x* runs through the positive integers. This is known as Dirichlet's theorem on the existence of primes in arithmetic progression, and we make use of this important result in a later chapter of the paper.

1.2: Alternative Methods for Counting Primes

In the absence of proofs of any of these open problems, recent research has used less direct methods and *heuristic* evidence to support many of the early conjectures concerning prime numbers. In this paper, we investigate some of these less direct methods, concentrating on the theory of *primitive divisors* , their properties, and how they are used as evidence of an infinitude of primes in particular sequences.

Definition. A prime number *p* is called a *primitive divisor* of a term P_n of a sequence *P* if *p* divides P_n but does not divide any other term of the sequence prior to P_n.

We begin our analysis of some of these "open problems" in Chapter 2 as we discuss the *Mersenne* sequence. This sequence has been the subject of much study over the past few hundred years as it is conjectured to have an infinite number of prime terms. We show how the Prime Number Theorem can be used to support this conjecture, and how the Mersenne sequence is a special case of a class of linear recurrence sequences which were shown by Zsigmondy to have an infinite number of terms with primitive divisors.

Sequences of the form studied by Zsigmondy are *binary linear recurrence sequences*, and in Chapter 3 we examine sequences of this type in a more general form, providing some examples to illustrate necessary conditions on the factors of the characteristic polynomials of these sequences for them to contain terms with an infinite number of primitive divisors. We also discuss *bilinear recurrence sequences* in this chapter, examining a famous example known as the *Somos-4* sequence.

Finally, in Chapter 4, we turn our attention to quadratic polynomial sequences of the form $n^2 + \beta$ with $-\beta$ not a square. This sequence has long been suspected to contain an infinite number of prime terms, however no proof is known for even one value of β. We instead examine the apparently simpler question as to whether the terms of the sequence generate an infinite number of primitive divisors. Everest, Stevens, Tamsett, and Ward have conducted extensive research on this subject in the past few years, and we describe much of their work in detail here, with an inclusion without proof of their most recent theorem which states that the number of terms of sequences of the form $n^2 + \beta$ with $-\beta$ having a primitive divisor has a natural density. We also find a relation between the number of primitive divisors of the sequence $n^2 + \beta$ and the number of lopsided integers in the sequence.

CHAPTER 2: THE MERSENNE SEQUENCE

2.1: Introduction to the Mersenne Sequence

One of the most notable unsolved problems remaining from the early study of elementary number theory, the Mersenne Prime Conjecture, involves the sequence $M = (M_n)_{n\geq 1}$, where

$$M_n = 2^n - 1. \tag{2-1}$$

Sequence (2-1) is known as the *Mersenne sequence*, named after the French mathematician Marin Mersenne (1588-1648). It is well known that if $2^n - 1$ is prime then n is prime for $n = 1, 2, \ldots$, however the converse is not always true. For example, if $n = 11$ then $2^n - 1 = 2047 = 23 * 89$.

Definition 2.1. If for a positive integer n, $M_n = 2^n - 1$ is prime, then M_n is called a *Mersenne prime*.

The Mersenne Prime Conjecture states that there are an infinite number of prime terms in the sequence M. Calculations show that the first several terms of the sequence are

$$M_1 = 1,\ M_2 = 3,\ M_3 = 7,\ M_4 = 15,\ M_5 = 31,\ M_6 = 63,\ M_7 = 127, \ldots, \tag{2-2}$$

and indeed M_n is prime for $n = 2, 3, 5,$ and 7. Mersenne conjectured that M_n was prime for $n = 2, 3, 5, 7, 13, 17, 19, 31, 67, 127,$ and 257 and composite for any other $n < 257$. While his guesses were somewhat inaccurate (he missed $n = 61, 89,$ and 107, while 67 and 257 were later shown to be composite), he provided a useful framework that would be used for many years to come.

Closely related to the Mersenne primes are the *perfect numbers*. A perfect number is a positive number *n* such that the sum of all of its proper divisors is equal to *n* itself, such as $6 = 1 + 2 + 3$. Another example of a perfect number is 28, since $28 = 1 + 2 + 4 + 7 + 14$. Now, note that $6 = 2^{2-1}(2^2 - 1)$, and that $2^2 - 1$ is the first Mersenne prime. Likewise, $28 = 2^{3-1}(2^3 - 1)$ and $2^3 - 1$ is the second Mersenne prime. In fact, Euclid showed that a number of the form $2^{n-1}(2^n - 1)$,where *n* and $2^n - 1$ are prime, is an even perfect number. (see [1, Chap.4]). Two thousand years later, Euler proved the converse of Euclid's theorem. It is not known whether any odd perfect numbers exist. The Mersenne Prime Conjecture could just as well be known as the Perfect Number Conjecture, since the search for the number of Mersenne primes is also the search for even perfect numbers.

While the first seven terms of the Mersenne sequence contain 3 primes, the frequency at which later primes appear quickly diminishes. As a matter of fact, only 43 Mersenne primes are known as of this writing. The 43rd was discovered on December 15, 2005 by Dr. Curtis Cooper and Dr. Steven Boone in conjunction with the Great Internet Mersenne Prime Search (GIMPS) [see www.mersenne.org]. This project, along with most other research involving the testing of possible Mersenne primes, uses a distributed version of the Lucas-Lehmer test for primality(see [3]). The algorithm itself is quite simple, however it involves doing extremely high-precision arithmetic which, even on today's

supercomputers, can take up to a month or more to check even one term of the Mersenne sequence. Hence, we are left with discoveries of new large primes from time to time, but as yet no rigorous proof of the Mersenne Prime Conjecture is known.

In the absence of such a proof, several have argued that there are infinitely many Mersenne primes using other less direct methods.

2.2: The Prime Number Theorem and Mersenne Primes

Definition. The *prime counting function,* $\pi(x)$, is the function giving the number of primes less than or equal to *x*. [4]

Thus, with $x = 10$, for example, $\pi(x) = 4$ since 2, 3, 5, and 7 are the only primes less than 10.

The Prime Number Theorem [17] states that

$$\pi(x) \sim \frac{x}{\log x}, \tag{2-3}$$

a result which we shall use often in this thesis. This implies that the probability of a number *n* chosen at random being prime is approximately $1/\log n$, and a corollary is that if p_x is the *x*th prime, then

$$p_x \sim x \log x. \tag{2-4}$$

(See [5] for a proof.)

Using the Prime Number Theorem, we can argue in favor of the Mersenne Prime Conjecture as follows. The number of Mersenne primes M_n with $n < N$ is expected to be approximately

$$\sum_{n<N} \frac{1}{\log(2^n - 1)}, \tag{2-5}$$

as long as M_n is no more or less likely to be prime than any other number of similar size chosen at random. It is clear that for large values of *n*, this sum is approximately

$$\sum_{n<N} \frac{1}{\log 2^n}$$
$$= \frac{1}{\log 2} \sum_{n<N} \frac{1}{n}, \tag{2-6}$$

which is a constant multiplied by the Harmonic Series. Since this series is divergent, we are left with the suggestion that there are infinitely many Mersenne primes. One can take this heuristic argument a step further by noting the following:

$$\frac{1}{\log 2} \sum_{n<N} \frac{1}{n} \sim \frac{\log N}{\log 2}. \tag{2-7}$$

It would seem natural to argue that this is the number of Mersenne primes M_n, with $n < N$, however no one has been able to provide a proof as such. Still, Wagstaff [6] and Pomerance and Lenstra [7] looked for a refinement of this heuristic argument that would more closely match the available evidence. Their conjecture was that the number of Mersenne primes M_n with $n < N$ is asymptotically equal to

$$\frac{e^{\gamma}}{\log 2}\log N, \tag{2-8}$$

where γ is the Euler-Mascheroni constant defined by

$$\gamma = \lim_{n\to\infty}\sum_{i=1}^{n}\left[\frac{1}{i} - \log\left(1+\frac{1}{i}\right)\right]. \tag{2-9}$$

2.3: Primitive Divisors and Mersenne Sequences

Since no rigorous proof of the existence of infinitely many Mersenne primes is known, we now examine the apparently simpler question of existence of *primitive divisors* in the Mersenne sequence. Recalling the definition of primitive divisors from Chapter 1, it is very easy to calculate the terms of a particular sequence which have or do not have primitive divisor(s). Table 1 lists the first fifteen terms of the Mersenne sequence, the prime factorization of each term, and the primitive divisors of each term.

Table 1: Primitive Divisors of the First 15 Terms of the Mersenne Sequence

n	M_n	***prime factorization***	***prime factors***	***primitive divisor(s)***
1	1	-	-	-
2	3	3	3	3
3	7	7	7	7
4	15	$3*5$	3,5	5
5	31	31	31	31
6	63	3^2*7	3,7	-
7	127	127	127	127
8	255	$3*5*17$	3,5,17	17
9	511	$7*73$	7,73	73
10	1023	$3*11*31$	3,11,31	11
11	2047	$23*89$	23,89	23,89
12	4095	$3^2*5*7*13$	3,5,7,13	13
13	8191	8191	8191	8191
14	16383	$3*43*127$	3,43,127	43
15	32767	$7*31*151$	7,31,151	151

It is clear from the table that each term M_n has at least one primitive divisor with the exception of $n=1$ and $n=6$. In 1892, Zsigmondy showed that for each $n>6$, M_n has at least one primitive divisor [7]. Subsequently, he proved a similar result for sequences of the more general form a^n-b^n, known as Zsigmondy's Theorem.

Theorem 2.1[Zsigmondy] If $1 \le b \le a$ with *a* and *b* coprime, then $a^n - b^n$ has at least one primitive divisor with the following two exceptions:

1. $2^6 - 1^6$
2. $n = 2$ and $a + b$ is a power of 2.

He also proved that sequences of the form $a^n + b^n$, $a > b \ge 1$, have a primitive divisor for each term with the exception of $2^3 + 1^3 = 9$.

It is obvious that each term of the Mersenne sequence is of the form $a^n - b^n$ with $a = 2$ and $b = 1$. And while Zsigmondy's result is much less restrictive and weaker than the Mersenne Prime Conjecture, his work spurred a great deal of interest and further research into the arithmetic of similar sequences. Schinzel[12], for example, was able to extend Zsigmondy's work by establishing a large class of indices for which the Mersenne sequence M_n has a *composite primitive divisor* (meaning M_n has more than one primitive divisor). Among his results was a proof of the fact that M_{4k} has a composite primitive divisor for all odd $k > 5$.

CHAPTER 3 : LINEAR AND BILINEAR RECURRENCE SEQUENCES

3.1: Linear Recurrence Sequences

Definition. A sequence of numbers is called a *binary linear recurrence sequence* if it satisfies a linear recurrence relation of the form

$$U_{n+2} = cU_{n+1} + dU_n \ , \tag{3-1}$$

for all $n \geq 1$ and constants *c* and *d*.

Note that with $c = a + b$ and $d = -ab$, Zsigmondy's sequence $U_n = a^n - b^n$ discussed in Chapter 2 satisfies this relation since

$$\begin{aligned} a^{n+2} - b^{n+2} &= (a+b)\left[a^{n+1} - b^{n+1}\right] + (-ab)\left[a^n - b^n\right] \\ &= aa^{n+1} - ab^{n+1} + a^{n+1}b - bb^{n+1} - aba^n + abb^n \\ &= a^{n+2} - b^{n+2}. \end{aligned} \tag{3-2}$$

We now examine later work beyond Zsigmondy's, which has shown that it is not necessary that *a* and *b* be integers in Zsigmondy's Theorem.

Definition. Let P and Q be coprime integers and α and β be distinct roots of the equation

$$x^2 - Px - Q = 0 . \tag{3-3}$$

(In other words, α and β are the zeros of a monic irreducible quadratic polynomial with integer coefficients.) Then the *Lucas numbers* are the numbers *u* and *v* defined by

$$u_n = \frac{\alpha^n - \beta^n}{\alpha - \beta} \text{ and } v_n = \alpha^n + \beta^n , \tag{3-4}$$

and the Lucas sequences are given by $U = u_n$ and $V = v_n$. The *Lehmer numbers* are the numbers *u* defined by

$$u_n = \begin{cases} \dfrac{\alpha^n - \beta^n}{\alpha - \beta} & \text{if } n \text{ is odd} \\ \dfrac{\alpha^n - \beta^n}{\alpha^2 - \beta^2} & \text{if } n \text{ is even} \end{cases} , \tag{3-5}$$

with the Lehmer sequence given by $U = u_n$. Note that both Lucas and Lehmer sequences satisfy a binary linear recurrence relation. These sequences are widely used today in the testing of the primality of large numbers.

Carmichael, Ward, and Durst [9] proved that if α and β are both *real* numbers, then there exist primitive divisors of each term in a Lucas or Lehmer sequence for any $n > 12$.

As an example, consider the Fibonacci sequence F_n = 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, 233,… . This sequence has

$$F_n = u_n\left(\frac{1+\sqrt{5}}{2}, \frac{1-\sqrt{5}}{2}\right), \tag{3-6}$$

and calculations show that F_n does not have a primitive divisor for $n = 1, 2, 6,$ or 12. Extending these results still further, Schinzel and Stewart [12] considered α and β complex and found upper bounds for the *n* for which the *n*th term of the sequence does not have a primitive divisor. Improving their results, Voutier [10] reduced the upper bound to *n* = 30030 and this number has recently been reduced to *n* = 30 by Voutier and his fellow researchers Bilu and Hanrot [10]. It is striking that such a low upper bound has been discovered, as this allows us to quickly check any sequence of the form (3-4) or (3-5) by simply calculating the first 30 terms.

We now examine primitive divisors of linear recurrence sequences in a more general setting.

Definition. The sequence $L = \left(L_n\right)_{n\geq 1}$ is called a *linear recurrence sequence of order k* if it satisfies the relation

$$L_{n+k} = c_{k-1}L_{n+k-1} + \ldots + c_0 L_n, \tag{3-7}$$

for all $n \geq 1$ and no shorter relation. When $k = 3$ the sequence *L* is called a *ternary linear recurrence sequence* and when *k* = 4 it is known as a *quaternary linear recurrence sequence*.

Let α and β denote the zeros of a monic irreducible quadratic polynomial as in (3-3). Then the integer sequence $W(\alpha,\beta) = (W_n(\alpha,\beta))_{n\geq 1}$ defined by

$$W_n(\alpha,\beta) = (\alpha^n - 1)(\beta^n - 1) \tag{3-8}$$

is always a linear recurrence sequence. Now consider the following examples:

Example 3.1. [15] The sequence $B = -W(2+\sqrt{3}, 2-\sqrt{3})$ is given by

2, 12, 50, 192, 722, 2 700, 10 082, 37 632, 140 450, 524 172, 1 956 242, … (3-9)

and is a ternary sequence which satisfies the relation

$$B_{n+3} = 5B_{n+2} - 5B_{n+1} + B_n . \tag{3-10}$$

Although the fourth and sixth terms of *B* do not have primitive divisors, Everest, Stevens, Tamsett, and Ward[15] conjectured that all the terms of B_n, $n > 6$, do have at least one primitive divisor.

On the other hand, the following example illustrates that this is not always the case.

Example 3.2. [15] The sequence $C = -W\left(1+\sqrt{2}, 1-\sqrt{2}\right)$ is given by

$$2, 4, 14, 32, 82, 196, 478, 1\,152, 2\,786, 6\,724, 16\,238, \ldots, \tag{3-11}$$

and is a quaternary sequence satisfying

$$C_{n+4} = 2C_{n+3} + 2C_{n+2} - 2C_{n+1} - C_n . \tag{3-12}$$

However, the terms C_{2k}, for odd *k*, do not have primitive divisors. The reason for this is shown in the following theorem.

Theorem 3.3. If $\alpha\beta = -1$ and *k* is odd, $2k = 4l + 2$, then $W_{2k}(\alpha, \beta) = -\left[W_k(\alpha, \beta)\right]^2$.

Proof:

$$\begin{aligned} W_{2k}(\alpha,\beta) &= \left(\alpha^{2k}-1\right)\left(\beta^{2k}-1\right) = \left(\alpha^{4l+2}-1\right)\left(\beta^{4l+2}-1\right) \\ &= \left(\alpha^{2l+1}-1\right)\left(\beta^{2l+1}-1\right)\left(\alpha^{2l+1}+1\right)\left(\beta^{2l+1}+1\right) \\ &= W_{2l+1}\left(\alpha^{2l+1}+1\right)\left(\beta^{2l+1}+1\right) \\ &= -W_{2l+1}\left[\left(\alpha^{2l+1}-1\right)\left(\beta^{2l+1}-1\right)\right] \\ &= -\left[W_k(\alpha,\beta)\right]^2 . \end{aligned} \tag{3-13}$$

Hence in *C* for example, $C_2 = C_1^2, C_6 = C_3^2$, etc, and in any sequence of the form *W*, the W_{2k} term will have the same prime factors as the W_k term and hence the sequence will contain an infinite number of terms without primitive divisors.

These are but two of many possible examples of sequences which can be analyzed in a general manner, as shown by Everest, et al. [15] Let

$$f(x) = x^k - c_{k-1}x^{k-1} - ... - c_0 \tag{3-14}$$

be the *characteristic polynomial* of (3-7). Then factoring *f* over the complex field yields

$$f(x) = (x-\alpha_1)^{e_1}...(x-\alpha_d)^{e_d}, \tag{3-15}$$

where the algebraic numbers $\alpha_1,...,\alpha_d$ are the *characteristic roots* of the sequence. Then the term L_n of any sequence satisfying (3-7) can be written as

$$L_n = \sum_{i=1}^{d} g_i(n)\alpha_i^n, \tag{3-16}$$

for polynomials $g_1, g_2,..., g_d$ of degrees $e_1 - 1, e_2 - 1,..., e_d - 1$ with algebraic coefficients.

Now consider the roots of the two sequences in the examples 3.1 and 3.2 above. When $\alpha\beta = 1$, as in 3.1, $W(\alpha,\beta)$ is a ternary linear recurrence sequence and has roots $1, \alpha,$ and β. However, when $\alpha\beta = -1$, as in 3.2, $W(\alpha,\beta)$ is a quaternary linear recurrence sequence with roots $1, -1, \alpha,$ and β. Everest conjectured that the terms of an integral linear recurrence sequence of order $k \geq 2$ will have a primitive divisor from some point on provided the roots are distinct and no pairwise quotient of distict roots is a

root of unity. This will be examined further in Chapter 4 as we investigate polynomial sequences.

3.2: Bilinear Recurrence Sequences

Definition. A sequence of numbers is called a *bilinear recurrence sequence* if it satisfies the bilinear recurrence relation

$$S_{n+4}S_n = S_{n+3}S_{n+1} + S_{n+2}^2, \text{ for all } n \geq 1, \text{ with } S_1 = S_2 = S_3 = S_4 = 1. \qquad (3\text{-}17)$$

The Somos-4 sequence is perhaps the most famous bilinear recurrence sequence with some fascinating properties. It is given by

$$1, 1, 1, 1, 2, 3, 7, 23, 59, 314, 1\,529, 8\,209, 83\,313, 620\,297, 7\,869\,898, \ldots . \qquad (3\text{-}18)$$

Note that the calculation of each term in the sequence involves division by S_n, resulting in a *rational* expression for the result. For example, to compute the next term in the sequence above, S_{16}, we have

$$S_{16} = \frac{S_{15}S_{13} + S_{14}^2}{S_{12}}$$

$$= \frac{(7\ 869\ 898 * 83\ 313) + (620\ 297)^2}{8\ 209}$$

$$= 126\ 742\ 987\,. \qquad \text{(3-19)}$$

Note that even though this computation involved division by S_n, the result is an integer. As a matter of fact, each term in the Somos-4 sequence is an integer. This sequence was discovered by Michael Somos [11] and is known to be associated with the arithmetic of elliptic curves. Propp [13] observed the fact that each term S_n is an integer is not so surprising, since the terms count matchings in a sequence of graphs.

Evaluating the first several terms of the Somos-4 sequence reveals that S_n is prime for $n = 5, 6, 7, 8, 9, 12, 14, \ldots$, and a natural question is whether there exist infinitely many primes in the sequence. Somos-4 can be further generalized to integer sequences *S* which satisfy a relation

$$S_{n+4}S_n = cS_{n+3}S_{n+1} + dS_{n+2}^2\,, \qquad \text{(3-20)}$$

where *c* and *d* are integer constants not both zero. Sequences satisfying this relation are generally called *Somos sequences*. Since a binary linear recurrence sequence will always satisfy (3-20), we define a Somos sequence to be *nonlinear* if it does not satisfy any linear recurrence relation. It seems natural to conjecture, since sequences of this type are generalizations of linear recurrence sequences, that they do contain an infinite number of

primes. However, heuristic evidence resulting from computations involving nonlinear Somos sequences has suggested otherwise. Results about the height of elliptic curves indicate a quadratic-exponential growth rate of S_n. Thus,

$$S_n \sim e^{hn^2}, \tag{3-21}$$

where h is some positive constant. Applying the Prime Number Theorem, the expected number of primes with $n < N$ should be

$$\begin{aligned} \sum_{n<N} \frac{1}{\log S_n} &\sim \sum_{n<N} \frac{1}{\log e^{hn^2}} \\ &= \frac{1}{h} \sum_{n<N} \frac{1}{n^2} \\ &\leq \frac{\pi^2}{6h} \text{ for all } N, \end{aligned} \tag{3-22}$$

which suggests that the number of primes in a Somos sequence is finite.

On the other hand, Silverman[5] was able to relate Zsigmondy's Theorem to sequences similar to the Somos-4 sequence through the study of elliptic curves, applying an analogue of the theorem which applies to Somos-4. . (See [6] for background on elliptic curves.) His lower bound was improved upon by Everest, McLaren, and Ward[4], who proved that for $n \geq 5$, each term S_n has at least one primitive divisor. Even so, a direct proof is yet to be found which applies to all Somos sequences, and in fact no more than a dozen prime terms have been discovered in any such sequence.

CHAPTER 4 : QUADRATIC POLYNOMIAL SEQUENCES

4.1: The Prime Number Theorem and Quadratic Polynomial Sequences

We now consider the sequence P defined by

$$P_n = n^2 + \beta, \tag{4-1}$$

an integer sequence defined by a monic quadratic polynomial. Any sequence of this type satisfies the linear recurrence relation

$$P_{n+3} = 3P_{n+2} - 3P_{n+1} + P_n \quad \text{for all } n \geq 1. \tag{4-2}$$

Mathematicians have long suspected that for any fixed integer β with $-\beta$ not a square, the sequence P will contain infinitely many prime terms. However, no proof is known today even for one value of β.

Recall that the Prime Number Theorem predicts that the sequence P defined by

$$P_n = n^2 + \beta \tag{4-3}$$

will have approximately

$$\sum_{n<N} \frac{1}{\log(n^2+\beta)} \tag{4-4}$$

prime terms, assuming that P_n is no more or less likely to be prime than any other random number of the same size. We examine this sum more closely and take the sum from $n=2$, without loss of generality, seeking an upper bound at first. We have

$$\begin{aligned}
\sum_{n=2}^{N} \frac{1}{\log(n^2+\beta)} &< \sum_{n=2}^{N} \frac{1}{\log n^2} \\
&= \frac{1}{2}\sum_{n=2}^{N} \frac{1}{\log n} \\
&< \frac{1}{2}\int_{2}^{N} \frac{dx}{\log x} \\
&= \frac{1}{2}\int_{2}^{N} \frac{x^{\varepsilon}dx}{x^{\varepsilon}\log x}\ ,\ (0<\varepsilon<1) \\
&< \frac{1}{2}\frac{N^{\varepsilon}}{\log N}\int_{2}^{N} x^{-\varepsilon}dx \\
&= \frac{1}{2}\frac{N^{\varepsilon}}{\log N}\left(N^{-\varepsilon+1}-2^{-\varepsilon+1}\right) \\
&\sim \frac{N}{2\log N} \to \infty \ \text{ as } \ N \to \infty .
\end{aligned} \tag{4-5}$$

By a similar argument, one can find an identical upper bound for the sum which indicates the asymptotic equality. This result seems to support the theory that there exist an infinite number of prime terms in the sequence *P*. Computation suggests that for fixed β there are

$$\frac{cN}{\log N} \tag{4-6}$$

prime terms with $n < N$, where $c = c(\beta)$ is some constant which depends on β. Bateman and Horn [2] provided a heuristic argument and numerical evidence to show that

$$c = \frac{1}{2}\prod_p \left(1-\frac{1}{p}\right)^{-1}\left(1-\frac{\omega(p)}{p}\right), \tag{4-7}$$

where the product is taken over all primes and $\omega(p)$ is the number of solutions *x* (mod *p*) to the congruence

$$x^2 \equiv -\beta \pmod{p}. \tag{4-8}$$

As in the previous sections, let us now consider the apparently simpler question as to whether the terms of the sequence contain primitive divisors. Of course, the terms which are prime are themselves primitive divisors. But what about the composite terms? Our

goal now is to determine the density of the terms of the sequence which have primitive divisors, and to that end we first examine a theorem only recently proven by Everest, Stevens, Tamsett, and Ward.

4.2: Lopsided Numbers

We are primarily concerned with asymptotic behavior and thus let us assume from this point on that $n > |\beta|$. Any term P_n with $n \le |\beta|$ is not guaranteed to behave as those described in the next sections.

Theorem 4.1. For all $n > |\beta|$, p is a primitive divisor of the term P_n if and only if p divides P_n and $p > 2n$.

Proof. Consider a prime $p < n$ which divides P_n. Then $P_n \equiv 0 \pmod{p}$, and thus we can have $P_m \equiv 0 \pmod{p}$ if we choose some $m < p$ where m is the residue of $n \pmod{p}$. Now $p < n$ and $m < p$, thus $m < n$. Hence p cannot be less than n and be a primitive divisor of P_n. Furthermore, since a solution to $P_m \equiv 0 \pmod{p}$ is guaranteed for some $m \le p/2$, then if p is a primitive divisor of the term P_n, $p > 2n$.

Conversely, assume that p is a prime which divides P_n but p is not a primitive divisor of P_n. Then $n^2 + \beta \equiv 0 \pmod{p}$, and there is an integer $m < n$ with

$m^2 + \beta \equiv 0 (\text{mod } p)$. Hence, $m^2 - n^2 \equiv 0 (\text{mod } p)$, and it is thus clear that $m \pm n \equiv 0 (\text{mod } p)$. In particular, $p \le m + n < 2n$.

It follows that a prime *p* is a primitive divisor of P_n if and only if *p* divides P_n and $p > 2n$.

Definition 4.1. An integer *k* is called *lopsided* if it has a prime factor *q* with $q > 2\sqrt{k}$.

Obviously any prime greater than 3 is lopsided. 12, 14, 52, and 69 are examples of non-lopsided numbers. Since the next few sections discuss their distribution, it is helpful to note that the first 30 terms of the sequence L_n of lopsided numbers is:

5, 7, 11, 13, 17, 19, 22, 23, 26, 29, 31, 34, 37, 38, 39, 41, 43, 46, 47, 51, 53, 57, 58, 59, 61, 62, 67, 68, 69, 71,…

Everest, et al. gave the following two important results:

Theorem 4.2. For all $n > |\beta|$, P_n has a primitive divisor if and only if P_n is lopsided.

Proof. First of all, note the requirement that $n > |\beta|$. This requirement is necessary, since if $|\beta|$ is prime then $P_{|\beta|}$ has a primitive divisor but is not lopsided. Now assume that *p* is a primitive divisor P_n. By theorem 4.1, $p > 2n$. Thus

$$p \ge 2n + 1 > 2\sqrt{n^2 + n} > 2\sqrt{n^2 + \beta} \tag{4-9}$$

and hence P_n is lopsided.

Conversely, let P_n be lopsided. Then there exists a prime *p* such that

$$p > 2\sqrt{n^2+\beta} \geq 2\sqrt{n^2-n+1} > 2n-1. \tag{4-10}$$

Since 2*n* cannot be prime, then $p > 2n$. Further, since *p* divides P_n and $p > 2\sqrt{P_n}$, then no other prime divisor can be as large, and thus *p* is a primitive divisor of P_n.

Theorem 4.3. Suppose $-\beta$ is not a square integer. Then there exist infinitely many terms of the sequence *P* which do not have a primitive divisor.

Proof. Schinzel [12] proved that for any positive *x*, the largest prime factor of $n^2+\beta$ is bounded above by n^x for infinitely many *n*. Choosing $x=1$, it is clear that $n^2+\beta$ is not lopsided infinitely often.

We now consider quantitative information about the frequency with which, rather than the extent to which, the terms of *P* have or do not have a primitive divisor. Based on the statement of Theorem 4.2, Everest suggested a stronger result of Theorem 4.3:

Conjecture 4.4. Suppose $-\beta$ is not a square and let $\rho_\beta(N)$ denote the number of terms P_n, with $n < N$, having a primitive divisor in the sequence *P*. Then

$$\rho_\beta(N) \sim cN, \tag{4-11}$$

for some constant $0 < c < 1$.

Although Everest was unable to find a proof of Conjecture 4.4, he found Theorem 4.3 helpful in providing heuristic evidence in its support. While Theorems 4.1 and 4.3 give some insight into the arithmetic of the sequence *P*, they do not give any suggestions for how often primitive divisors should occur. In the last section of this chapter, we mention some approaches to bounding the number of terms in *P* which have primitive divisors. First, we note that an asymptotic formula can be obtained for the distribution of lopsided numbers.

4.3: Distribution of Lopsided Numbers

Theorem 4.5. [Everest] Let $\pi_i(N)$ denote the number of lopsided numbers less than or equal to *N*. Then

$$\pi_i(N) \sim N \log 2 \text{ as } n \to \infty. \tag{4-12}$$

Proof. Let *k* be an arbitrary lopsided number and write $k = pm$ where *p* is the largest prime factor of *k*. Since *k* is lopsided, then $p > 2\sqrt{pm} \Rightarrow p^2 > 4pm \Rightarrow p > 4m$. We will count the number of lopsided integers less than *N* by partitioning them into two groups. First, let *p* be a variable prime. For $p < 2\sqrt{N}$ there are $\left\lfloor p/4 \right\rfloor$ lopsided integers $pm < N$. For $p \geq 2\sqrt{N}$, all *pm* such that $pm < N$ are lopsided and hence there are $\left\lfloor N/p \right\rfloor$ lopsided integers in this group. Thus the total number of biased integers less than *N* is

$$\sum_{p<2\sqrt{N}} \left\lfloor \frac{p}{4} \right\rfloor + \sum_{2\sqrt{N} \le p<N} \left\lfloor \frac{N}{p} \right\rfloor. \tag{4-13}$$

Let us examine each of the terms in this sum separately. Letting

$$\sum\nolimits_1 = \sum_{p<2\sqrt{N}} \left\lfloor \frac{p}{4} \right\rfloor \text{ and } \sum\nolimits_2 = \sum_{2\sqrt{N} \le p<N} \left\lfloor \frac{N}{p} \right\rfloor, \tag{4-14}$$

we consider $\sum_1$ first. This sum is certainly less than the sum of all primes less than $2\sqrt{N}$ and by the Prime Number Theorem is thus $O\left(N/\log N\right)$. Since we are interested in asymptotic behavior, we will ignore this sum.

Next we consider $\sum_2$ and its behavior. Note that $\sum_2$ can be written as

$$\sum\nolimits_2 = \sum_{p=2\sqrt{N}}^{N-1} \left[\left\lfloor \frac{N}{p} \right\rfloor - \frac{N}{p} \right] + N \sum_{p=2\sqrt{N}}^{N-1} \frac{1}{p}, \tag{4-15}$$

and since the first term in this sum is $O\left(N/\log N\right)$ we are left to estimate the sum

$$N \sum_{p=2\sqrt{N}}^{N-1} \frac{1}{p}. \tag{4-16}$$

For this we apply *Merten's formula*[18],

$$\sum_{p<x} \frac{1}{p} = \log\log x + A + o(N)\,, \tag{4-17}$$

where A is known as *Merten's constant* and $A \approx 0.2614972...$. Applying this to (4-17) we have

$$N \sum_{p=2\sqrt{N}}^{N-1} \frac{1}{p} = N\left[\log\log N + A - \log\left(\log 2 + \log\sqrt{N}\right) - A\right] + o(1)$$

$$= N\left[\log\left(\frac{\log N}{\log 2 + \log\sqrt{N}}\right)\right] + o(1)\,. \tag{4-18}$$

Now, as $N \to \infty$,

$$\left(\frac{\log N}{\log 2 + \log\sqrt{N}}\right) \sim \frac{\log N}{\log\sqrt{N}} = \frac{\log N}{\frac{1}{2}\log N} = 2\,, \tag{4-19}$$

and since $\log x$ is a continuous function, then

$$\log\left(\frac{\log N}{\log 2 + \log\sqrt{N}}\right) \sim \log 2 \text{ as } N \to \infty\,. \tag{4-20}$$

Thus we have

$$\sum\nolimits_2 \sim N \log 2, \tag{4-21}$$

and hence our desired result

$$\pi_l(N) \sim N \log 2 \text{ as } N \to \infty. \tag{4-22}$$

Note the following application of Theorem 4.5 to Conjecture 4.4: The probability that a large integer is lopsided can be estimated as roughly $\log 2$. Thus, the number of values of $n^2 + \beta$ with $n < N$ that we can expect to be lopsided is asymptotically $N \log 2$. Indeed, computations reveal that the number of terms which are lopsided in $n^2 + \beta$ is asymptotically cN for some constant *c*. Computations with $|\beta| < 20$ suggest that while convergence is slow, *c* is reasonably close to $\log 2$.

4.4: Conclusions

We conclude by stating some estimates for the number of terms in *P* which do have primitive divisors. Let $\rho(N)$ denote the number of terms P_n in *P* with $n < N$ that have primitive divisors. Everest, et al. only recently gave a proof of the following important result.

Theorem 4.6. There is a constant $C > 0$ such that

$$\rho(N) < N - \frac{CN}{\log N} \tag{4-23}$$

holds for all sufficiently large *N*. There is a constant $D > 0$ such that

$$\frac{N}{2} - \frac{DN}{\log N} < \rho(N). \tag{4-24}$$

The proof of Theorem 4.6 can be found in [15] for interested readers, and uses the product

$$Q_N = \prod_{n=1}^{N} |P_n|, \tag{4-25}$$

since counting the number of prime divisors of Q_N which are greater than or equal to $2N$ will be the same as counting the number of primitive divisors of P_n. (Primes which divide Q_N but are less than $2N$ can be ignored asymptotically.)

This area of research is still relatively new and many opportunities for further study abound. Everest, et al.[15] conjectured a slightly stronger result than Theorem 4.6:

Conjecture 4.7. Suppose that $-\beta$ is not a square. If $\rho_\beta(N)$ denotes the number of terms P_n with $n < N$ that have primitive divisors in the sequence *P*, then

$$\rho_\beta(N) \sim cN \ , \tag{4-26}$$

for some constant c satisfying $0 < c < 1$.

Everest, et al. have been unable to provide a proof of this conjecture as yet, and this is one topic which would be an excellent choice for further research as it provides a sharper estimate for the density of primitive divisors of the terms of P.

LIST OF REFERENCES

1) Archibald, R. (1970). *An introduction to the theory of numbers*. Columbus, OH: Merrill.
2) Prachar, K., (1978). *Primzahlverteilung*. Berlin: Springer-Verlag.
3) Bressoud, D. (1989). *Factorization and Primality Testing*. New York: Springer-Verlag.
4) Shanks, D. (1993). *Solved and Unsolved Problems in Number Theory, 4th ed.* New York: Chelsea.
5) Silverman, J. H. (1988). Wieferich's criterion and the abc-conjecture, *J. Number Theory, 30(2), 226-237*.
6) Silverman, J.H. & Tate, J. (1992). *Rational Points on Elliptic Curves*. New York: Springer-Verlag.
7) Zsigmondy, K. (1892). Zur Theorie der Potenzreste, Monatsh. *Math.*, *3*, 265-284.
8) Tijdeman, R. (2002) Some Applications of Diophantine Approximation, *Number Theory for the Millennium III*, Natick, NH: A.K. Peters.
9) Durst, L.K. (1961). Exceptional real Lucas sequences, *Pacific Journal of Mathematics, 11*(*2*), 489-494.
10) Bilu, Y., Hanrot, G., & Voutier, P.M. (2001). *Existence* of primitive divisors of Lucas and Lehmer numbers, *Math., 539*, 75-122.
11) Somos, M. (1989). Problem 1470, *Crux Mathematicorum , 15*, 208.
12) Schinzel, A. (1962). On primitive prime factors of $a^n - b^n$, *Proc. Cambridge Philos. Soc.*, *58*, 555-562.
13) Everest, G., Miller, V., & Stephens, N. (2004). Primes generated by elliptic curves, *Proc. Amer. Math. Soc.*, *132*(*4*), 955-963.
14) Everest, G., McLaren, G., & Ward, T. (2004). Primitive divisors of elliptic divisibility sequences, *arXiv:math.NT/0409540*.
15) Everest, G., Stevens, S., Tamsett, D., & Ward, T. (2005). Primes generated by recurrence sequences, *arXiv:math.NT/0412079*.

16) Apostol, T. (1986). *Introduction to Analytic Number Theory*. New York: Springer-Verlag.
17) Hadamard, J. (1896). Sur la distribution des zeros de la function zeta et ses consequences arithmetiques, *Bull. Soc. Math. France, 24*, 199-220.
18) Wright, G. & Wright, E. (1979). *An Introduction to the Theory of Numbers*. Oxford: Clarendon Press.