LYNDON WORDS OF A SECOND-ORDER RECURRENCE

BOB BASTASZ

Abstract. The sequence of digits forming the least period of the Fibonacci sequence (mod m) is a Lyndon word. Besides (0,1), other starting sequences can form Lyndon words that have a length equal to the least period of the recurrence $d_{i+2} \equiv d_i + d_{i+1} \pmod{m}$. Let $S(p)$ be the set of all such starting sequences, where p is a prime. Properties of this set are described, including its cardinality, n, and the number, c, of different length Lyndon words formed by elements in $S(p)$. Considering the fraction of possible Lyndon words that are in $S(p)$ leads to the development of a parameter called the period index, λ, which is related to the reciprocal of the Pisano period and concisely defines the main properties of $S(p)$.

1. Introduction

Interest in periodicities associated with the Fibonacci sequence dates back to at least the 18th century. Dickson mentions that Lagrange showed Fibonacci-like terms are periodic for any modulus [3]. The topic has flourished over the last century and numerous studies have been published about the periodic behavior of sequences defined by a linear recurrence relation (e.g., [2, 12, 13, 4, 5, 11, 10]).

Recurrence relations can generate Lyndon words. A notable example is the sequence of digits that comprise the least period of the Fibonacci sequence (mod m). Here we consider various finite sequences of digits that follow a particular linear recurrence relation and are Lyndon words. After noting that a starting sequence which generates a Lyndon word must itself be a Lyndon word (with one exception), we evaluate the fraction of Lyndon word starting sequences that generate Lyndon words of a length equal to the least period of the recurrence. This motivates introducing an integer parameter called the period index, which quantifies a relationship between the recurrence and Lyndon words formed from it.

2. Lyndon words

Using one of several equivalent definitions [1], we define a Lyndon word to be a finite sequence of digits that is minimal among all its cyclic rotations. Specifically, $(d_1, d_2, \ldots, d_k)_p$ is a k-length Lyndon word if it is strictly less than any of the circular shifts $(d_i, \ldots, d_k, d_1, \ldots, d_{i-1})_p$ for $1<i\leq k$. Here "less than" means the smaller in value of two k-digit numbers to the base p indicated by the subscript. Lyndon words are primitive in the sense they cannot be a power of a smaller word. For example $(0, 2, 1)_3$ is a Lyndon word while $(0, 2, 1, 0, 2, 1)_3 = (0, 2, 1)_3^2$ is not.

3. Lyndon words generated by a recurrence

Lyndon words can be constructed using a recurrence relation. We will consider the second-order linear recurrence

$$d_{i+2} \equiv d_i + d_{i+1} \pmod{p}, \tag{3.1}$$

where p is a prime, with the starting sequence $(d_1, d_2)_p$. Some starting sequences generate a Lyndon word of the form $(d_1, d_2, \ldots, d_k)_p$, where k is the least period of the recurrence.

The length of a Lyndon word formed from the starting sequence $(d_1, d_2)_p$ will be denoted by $k(d_1, d_2)_p$.

Two examples are $(0, 0)_p$, which produces the Lyndon word $(0)_p$ of length 1, and $(0, 1)_p$, which produces a k-length Lyndon word where k is the least period of the Fibonacci sequence (mod p) (i.e., the Pisano period).

4. Starting sequences that form Lyndon words

It is interesting to consider which starting sequences $(d_1, d_2)_p$ form k-length Lyndon words that follow (3.1). For a given p, let $S(p)$ be the set of such starting sequences: $S(p) = \{(d_1, d_2)_p \mid$ (3.1) forms a k-length Lyndon word $(d_1, d_2, \ldots, d_k)_p$ with $k \geq 1\}$. In the following, we examine its cardinality, $n = |S(p)|$, and the number of distinct word lengths generated by the starting sequences in $S(p)$, denoted c. Clearly, starting sequences must themselves be Lyndon words, or a power of a Lyndon word if k is less than the order of the recurrence. So, for a second-order recurrence, only starting sequences that are length two Lyndon words, or are the second power of a length one Lyndon word, need be considered.

The number of length l Lyndon words that can be formed using m digits, denoted $\mathcal{L}(m, l)$, is given by

$$\mathcal{L}(m, l) = \frac{1}{l} \sum_{d|l} \mu(l/d)\, m^d, \tag{4.1}$$

where μ is the Möbius function [8]. For prime p, (4.1) gives $\mathcal{L}(p, 1) = p$ and $\mathcal{L}(p, 2) = p(p-1)/2$. Consequently, $n \leq \mathcal{L}(p, 1) + \mathcal{L}(p, 2) = p(p+1)/2$. Furthermore, if a starting sequence $(d_1, d_1)_p$ forms a length one Lyndon word, then $d_1 + d_1 \equiv d_1 \pmod{p}$ for $d_1 \in \mathbb{Z}_p$. This condition implies $d_1 = 0$, so only the $(0, 0)_p$ starting sequence forms a length one Lyndon word. Consequently $n \leq 1 + p(p-1)/2$.

We note that for all p, $\{(0, 0)_p, (0, 1)_p\} \subseteq S(p)$. No ordered pair $(d_1, d_2)_p$ has a value lower than $(0, 0)_p$ and since only $(0, 0)_p < (0, 1)_p$, both $(0, 0)_p$ and $(0, 1)_p$ form Lyndon words. $(0, 0)_p$ forms $(0)_p$ and $(0, 1)_p$ generates $(0, 1, \ldots, p\text{–}1, 1)_p$ with $k(0, 1)_p$ equal to the Pisano period. Because of their ubiquity, $(0, 0)_p$ and $(0, 1)_p$ are given their own symbols, z and u, respectively. Since $k(z) = 1$ and $k(u) \geq 2$, $n \geq 2$ and $c \geq 2$ for every $S(p)$.

5. Characteristic polynomial analysis

Associated with recurrence (3.1) is the characteristic polynomial $f(x) = x^2 - x - 1 \pmod{p}$. This is a monic, quadratic polynomial with $f(0) \neq 0$. Finite field theory [14, 7] provides a direct way to evaluate n and c for a given p depending on the properties of $f(x)$ in $\mathbb{F}_p[x]$.

There are four cases to consider. First, if $f(x)$ is irreducible, $k(u) = \mathrm{ord}(f(x))$, where $\mathrm{ord}(f(x))$ is the order of $f(x)$. Besides z, there are $(p^2 - 1)/k(u)$ elements in $S(p)$ that form words of length $k(u)$. Hence $n = 1 + (p^2 - 1)/k(u)$ and $c = 2$.

Next, if $f(x)$ reduces to a single, squared term such that $f(x) = (x + a)^2$ and $a \neq 0$, then $a^2 = 2a$ and $a = 2$. This implies $4 \equiv -1 \pmod{p}$ and therefore $p = 5$. By enumeration one finds $k(u)_5 = 20$ and $k(1, 3)_5 = 4$, so $S(5) = \{z, u, (1, 3)_5\}$, $n = 3$ and $c = 3$.

Finally, if $f(x)$ is reducible to distinct linear factors, $f_a(x)$ and $f_b(x)$, there are two possibilities. The two factors have equal orders in $\mathbb{F}_p[x]$ if the order is divisible by 4 ([7], theorem 3.14). Then $S(p)$ has the same structure as when $f(x)$ is irreducible, with $\mathrm{ord}(f_a(x))$ (or $\mathrm{ord}(f_b(x))$) replacing $\mathrm{ord}(f(x))$. So $k(u) = \mathrm{ord}(f_a(x))$, $n = 1 + (p^2 - 1)/k(u)$ and $c = 2$.

Now suppose the characteristic polynomial is reducible and its factors have different orders in $\mathbb{F}_p[x]$. The linear factors $f_a(x)$ and $f_b(x)$ form sets $S_a(p)$ and $S_b(p)$. Since each factor is irreducible, each set contains z and $(p - 1)/k'(u)$ elements that form sequences of length

$k'(u)$ where $k'(u)$ is the order of $f_a(x)$ or $f_b(x)$, as appropriate. Words are formed by taking all combinations of elements, one from each set, subject to the constraint of being a Lyndon word. In this case the orders of $f_a(x)$ and $f_b(x)$ in $\mathbb{F}_p[x]$ differ by a factor of two ([7], theorem 3.14), so $S(p)$ contains sequences that form words of lengths 1, $k(u)$, and $k(u)/2$. This yields $n = 1 + (p^2 + p - 2)/k(u)$ and $c = 3$.

The quadratic character of 5 (mod p) tells whether or not $f(x)$ has factors. Completing the square of $f(x)$ shows $(2x - 1)^2 \equiv 5 \pmod{p}$, so five must be a quadratic residue modulo p if $f(x)$ is reducible. The Legendre symbol $(\frac{p}{5})$ indicates whether or not this is the case. $(\frac{p}{5})$ is 1 when $p \equiv \pm 1 \pmod{10}$ and –1 when $p \equiv \pm 3 \pmod{10}$, from which it follows that $f(x)$ is reducible when the prime modulus (expressed in base 10) ends in 1 or 9 and is irreducible when p ends in 3 or 7.

To summarize, except for $p = 5$, the characteristic polynomial $x^2 - x - 1 \pmod{p}$ falls into one of three categories. It is irreducible if $p \equiv \pm 3 \pmod{10}$ and reducible if $p \equiv \pm 1 \pmod{10}$. If reducible, the orders of its factors may be either equal or unequal.

6. Lyndon word fraction and the period index λ

The set $S(p)$ is comprised of Lyndon words having lengths one or two. It's natural to ask, given p, what fraction of all possible Lyndon words of these lengths are members of $S(p)$? As shown above, every set $S(p)$ contains exactly one element that produces a word of length one, namely z. Consequently, the fraction of length 1 Lyndon words in $S(p)$ is $1/\mathcal{L}(p, 1)$, which from (4.1) is just $1/p$.

More interesting is the fraction of length 2 Lyndon words in $S(p)$. Applying (4.1), the fraction of length 2 Lyndon words in $S(p)$ is

$$\frac{n-1}{\mathcal{L}(p,2)} = \frac{2\,(n-1)}{p\,(p-1)}. \tag{6.1}$$

Let Γ represent this ratio. It's informative to look at Γ as a function of p ($p \neq 5$) for each type of $f(x)$. To simplify matters we introduce a parameter λ, which will be called the *period index*, defined as

$$\lambda = \frac{2\,(p-\varepsilon)}{k(u)}, \tag{6.2}$$

where ε represents the Legendre symbol $(\frac{p}{5})$. Values of λ are seen to be integers proportional to the reciprocal of $k(u)$.

When $f(x)$ is irreducible, $n - 1 = (p^2 - 1)/k(u)$, so $\Gamma = 2(p + 1)/(p\,k(u)) = \lambda/p$. λ is an integer because when $p \equiv \pm 3 \pmod{10}$, $k(u) \mid 2(p + 1)$ ([11], theorem 7; [10], theorem 2). Furthermore, $k(u) \nmid (p + 1)$, so $(2(p + 1)/\lambda) \nmid (p + 1)$ and $2 \nmid \lambda$. Consequently, when $f(x)$ is irreducible, λ is an odd integer.

When $f(x)$ is reducible, $\lambda = 2(p-1)/k(u)$. If $\operatorname{ord}(f_a(x)) = \operatorname{ord}(f_b(x))$, $n-1 = (p^2-1)/k(u)$ and $\Gamma = \lambda(p+1)/(p(p-1))$. Otherwise $n-1 = (p^2+p-2)/k(u)$ and $\Gamma = \lambda(p+2)/(p(p-1))$. Note that since $k(u)|(p - 1)$ ([11], theorem 6; [10], theorem 2), $2|\lambda$ and so, when $f(x)$ is reducible, λ is an even integer.

Combining the above expressions for Γ, we obtain a general relation for the fraction of length 2 Lyndon words as a function of λ and p. It is

TABLE 1. Prime sequences associated with λ

λ	c	OEIS entry	initial values
1	2	A071774	3,7,13,17,23, ...
2	2,3	A003147	(5),11,19,31,41,59, ...
3	2	A308784	47,107,113,263,347, ...
4	2,3	A047650	29,89,101,181,229, ...
5	—	—	—
6	3	A308796	139,151,331,619,811, ...
7	2	A308785	307,797,1483,3023,4157, ...
8	2,3	A308789	769,809,1049,1289,1721, ...
9	2	A308786	233,557,953,4013,4733, ...
$10i$	2,3	A001583	211,281,421,461,521, ...

i=1,2,3, ...

$$\Gamma = \frac{\lambda\,(p+\alpha)}{p\,(p-1)}, \tag{6.3}$$

where

$$\alpha = \begin{cases} -1, & \text{if } f(x) \text{ is irreducible,} \\ 1, & \text{if } f(x) \text{ is reducible and } \operatorname{ord}(f_a(x)) = \operatorname{ord}(f_b(x)), \\ 2, & \text{if } f(x) \text{ is reducible and } \operatorname{ord}(f_a(x)) \neq \operatorname{ord}(f_b(x)). \end{cases} \tag{6.4}$$

In the first case $\Gamma = \lambda/p$. In the other two cases Γ approaches λ/p as p increases. Except for a few small primes, $\Gamma \approx \lambda/p$ is a reasonable approximation.

The period index λ is a simple parameter for specifying the properties of $S(p)$ and for categorizing many of the sequences related to Fibonacci numbers. Primes with the same λ constitute various well-known sequences [9] as shown in Table 1 for λ=1-10. For example, the sequence of primes listed for $\lambda=10i$, where $i \in \mathbb{N}$, forms the artiads [6]. There appear to be no primes for which λ is an odd multiple of five.

7. The structure of $S(p)$

Using the results of the previous section, we can develop a general formulation for the structure of $S(p)$. For conciseness, we introduce two parameters, $\beta = (p-\varepsilon)/\lambda$ and $\omega = \varepsilon(1+(-1)^{\beta})/2$. The parity of β determines whether $S(p)$ has $c = 2$ or 3. Parameter ω controls the occurrence of Lyndon words that have length $k(u)/2$. Knowing λ is sufficient to determine the main properties of $S(p)$. Thus, given λ for $S(p)$,

$$n = 1 + \lambda\,(p+\omega)\,/\,2 + \lambda\,(1-\omega^2), \tag{7.1}$$

$$c = 3 - \omega^2. \tag{7.2}$$

An analysis of the characteristic polynomial or numerical calculation of $k(u)$ is needed to find the value of λ but, once obtained, λ can be used as a compact descriptor of $S(p)$ and may be considered to be a defining characteristic.

References

[1] J. Berstel and D. Perrin, The origin of combinatorics on words, *European J. Combin.* **28** (2007), 996–1022.

[2] R. D. Carmichael, On sequences of integers defined by recurrence relations, *Quart. J. Pure Appl. Math.* **48** (1920), 343–372.

[3] L. E. Dickson, *History of the Theory of Numbers, Volume 1*, Carnegie Institution of Washington, Washington, 1919. p. 393.

[4] H. T. Engstrom, On sequences defined by linear recurrence relations, *Trans. Amer. Mat. Soc.* **33** (1931), 210–218.

[5] M. Hall, An isomorphism between linear recurring sequences and algebraic rings, *Trans. Amer. Mat. Soc.* **44** (1938), 196–218.

[6] E. Lehmer, Artiads characterized, *J. Math. Anal. Appl.* **15** (1966), 118–131.

[7] R. Lidl and H. Niederreiter, *Introduction to Finite Fields and Their Applications*, Cambridge University Press, Cambridge, 1994.

[8] R. C. Lyndon, On Burnside's problem, *Trans. Amer. Math. Soc.* **77** (1954), 202–215.

[9] OEIS Foundation Inc., *The on-line encyclopedia of integer sequences*, (2011), `https://oeis.org`.

[10] A. Vince, Period of a linear recurrence, *Acta Arith.* **39** (1981), 303–311.

[11] D. D. Wall, Fibonacci series modulo m, *Amer. Math. Monthly* **67** (1960), 525–532.

[12] M. Ward, The characteristic number of a sequence of integers satisfying a linear recursion relation, *Trans. Amer. Mat. Soc.* **33** (1931), 153–165.

[13] M. Ward, The arithmetical theory of linear recurring series, *Trans. Amer. Mat. Soc.* **35** (1933), 600–628.

[14] N. Zierler, Linear recurring sequences, *J. Soc. Indust. Appl. Math.* **7** (1959), 31–48.

MSC2010: 11B01, 05A02, 60C02

P. O. Box 8286, Missoula, MT 59807
E-mail address: `bastasz@protonmail.com`