Ithaca College Mathematics Capstone

The Thue-Morse Sequence

- Definitions, Properties & Applications -

By Kenneth Potts

May 13, 2017

Contents

Bio of The Author

Kenneth Potts is a senior student of Mathematics at Ithaca College, graduating in May of 2017. He entered Ithaca College as a Television and Radio major, but his interest in math soon developed and took over his academic interests. After taking many mathematics courses, he developed an interest in data science, as well as computational, statistical, and financial mathematics, as well as computer programming. Apart from his mathematical interests, he loves playing guitar in his band, reading new books, and learning new concepts.

Abstract

The Thue-Morse Sequence is a very specific infinite binary sequence. The Sequence may be written using 0s and 1s as 0110100110010110... or as a word using As and Bs as $ABBABAABBAABABBA$.... There are multiple ways to define or derive the sequence. The sequence can be used as a fair sequence for taking turns, pouring coffee, and also as a way to partition 2^n integers into disjoint sets which sum to the same value. This is known as the Prouhet-Tarry-Escott problem. The sequence can also be derived as the solution to a generating quadratic equation. The sequence can generate another sequence which is square-free, having no adjacent repeating sections. The Thue-Morse sequence can be generalized to other bases and dimensions.

1 Introduction

I have researched sequences before, specifically one related to the Collatz Conjecture. It is an area that interests me a lot. The Thue-Morse sequence is an impartial and ubiquitous sequence. It appears to pop up all throughout mathematics from seemingly thin air. It can dictate the fair way to take turns, how multiple cups of coffee should be poured, and is the solution to multiple apparently unrelated problems in mathematics. It draws upon ideas from the fields of discrete mathematics, combinatorics, abstract algebra, linear algebra, and calculus. I have chosen to study this topic to strengthen and develop my understanding of these fields while simultaneously diving into a truly interesting topic. There are numerous ways to write the sequence, for example, using 0s and 1s as $T = 0110100110010110$... or as a word using As and Bs as $T = ABBABAABBAABABBA....$ The sequence was first created by a Norwegian mathematician named Axel Thue. He was interested in sequences with limited to no repetitions[1].

2 The Sequence Defined

The Thue-Morse sequence has many interesting properties. Its many distinct characteristics allow for multiple definitions based on those traits. The sequence will be referred to as T and as previous shown, the sequence begins as $T = 0110100110010110....$ By definition, T has no end. T is usually formed by iterating a certain set of rules. There are multiple rules that can be used to generate the sequence resulting in multiple definitions. To begin we will start with a formal definition.

2.1 The Bitwise Definition

We will begin by defining the n^{th} iteration of the formation of the sequence as T^n, where $T^0 = 0$. We will also define the complement of T^n as $\overline{T^n}$ where all the elements of T^n have been replaced with their bitwise negation to form $\overline{T^n}$[4, 3]. We can write the bitwise negation for an element $x \in \{0, 1\}$ as $\overline{x} = (1 - x) \bmod 2$. Therefore the bitwise negation of 0 is 1, and 1 is 0. This negation can also be written in terms of the Thue-Morse Sequence elements as $\overline{T_i} = (1 - T_i) \bmod 2$ performed on all the elements of T^n individually where we define the i^{th} element of the sequence T as T_i for $i \in \mathbb{N}$ [1]. For example, $T^1 = 01$, therefore $\overline{T^1} = 10$, as each element is individually negated. Then the n^{th} iteration of the sequence is given by appending $\overline{T^{n-1}}$ to T^{n-1}.

$$T^n = T^{n-1}\overline{T^{n-1}} \tag{1}$$

We can then form the iterations

$$T^0 \mapsto T^1 \mapsto T^2 \mapsto T^3 \mapsto T^4 \tag{2}$$

$$\implies 0 \mapsto 01 \mapsto 0110 \mapsto 01101001 \mapsto 0110100110010110 \tag{3}$$

Then we can define the full Thue-Morse sequence as [4]

$$\lim_{n\to\infty} T^n = T = 01101001100101101001011001101001... \tag{4}$$

Each new iteration doubles the number of elements in the sequence, and therefore there are 2^n elements in the n^{th} iteration of the sequence. Then by this definition of the sequence, any element T_i in the first 2^k elements is the negation of the $(i+2^k)^{th}$ element, for $k \in \mathbb{N}$. This property is an essential component of a coming proof. Also note that we can find the k^{th} iteration of the sequence as T^k where

$$T_0T_1T_2...T_{2^k-1} = T^k \tag{5}$$

2.2 The Recursive Definition

With the definition of T_i as the i^{th} element of the sequence T for $i \in \mathbb{N}$, element T_i can be defined recursively. Let $T_0 = 0$, then

$$T_{2n} = T_n \tag{6}$$

$$T_{2n+1} = \overline{T_n} \tag{7}$$

for all $n \geq 0$, where $T_i \in \{0, 1\}$ and again $\overline{T_n} = (1 - T_n) \bmod 2$[1].

2.3 The Binary Sum Definition

We will define the function $S_b(m)$ as the sum of the digits in the base-b expansion of the integer m. Then element T_i can be defined explicitly using $S_2(i)$ [1][6].

$$T_i = S_2(i) \bmod 2, i \geq 0 \tag{8}$$

As an example:

$$T_0 = S_2(0) \bmod 2 = 0 \tag{9}$$
$$T_1 = S_2(1) \bmod 2 = 1 \tag{10}$$
$$T_2 = S_2(2) \bmod 2 = 1 \tag{11}$$
$$T_3 = S_2(3) \bmod 2 = 0 \tag{12}$$

By the construction of the binary expansion of a number, we get $S_2(2n) = S_2(n)$ [1]. As an example, the number 12 has the binary representation 1100, and $2(12) = 24$ has the binary representation 11000. A single 0 has been added, which does not change the sum. Therefore $S_2(2n) = S_2(n)$. It also follows that $S_2(2n+1) = S_2(n) + 1$, since adding a 1 to an even number replaces the last 0 with a 1 in the binary representation[7]. As an example 1101 is the binary representation of 13.

2.4 The Substitution Definition

The Thue-Morse sequence can also be produced using an iterative substitution rule. The substitution rule can be defined by a homomorphism (mapping) $\tau : \{0,1\} \to \{01,10\}$. Let τ be defined by $0 \mapsto 01$ and $1 \mapsto 10$ [4][1]. Then iterating the morphism τ on the initial element 0 gives

$$0 \mapsto 01 \mapsto 0110 \mapsto 01101001 \mapsto 0110100110010110 \tag{13}$$

which is equivalent to mapping (3) given by the bitwise definition. Then iterating on τ infinitely for the initial element 0 gives the Thue-Morse sequence.

3 Equivalence of Constructions

The definitions presented are all ways to generate the sequence. It is necessary to confirm that the sequences produced by each definition are in fact equivalent. This can be shown by proving that although the definitions are different, the process which generates the sequence is the same.

3.1 Bitwise and Substitution Definitions

We will show that the bitwise and substitution definitions for the TM sequence are equivalent and yield the same sequence by constructing equivalent binary trees from both definitions. We will call this tree the TM tree. The TM tree has the property that the numbers reading across the n^{th} level of the tree are the ordered elements of the n^{th} iteration of the Thue-Morse sequence. Since the rules for both definitions are used to produce the same TM tree, the definitions are equivalent, and therefore yield the same sequence.

Proof. We will prove by induction that the definitions produce the same TM tree, are equivalent, and therefore yield the same sequence. We will construct the first iteration of the TM tree and therefore the first iteration of the TM sequence using the substitution definition in section 2.4. We have that $T^0 = 0$, and we will use the homomorphism τ to map from $T^0 = 0$ to $T^1 = 01$ as $\tau(0) = 01$. We can consider the TM tree for T^0 just the single element 0.

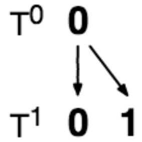

Figure 1: TM tree to the T^1 level via the substitution definition

The TM tree is constructed as shown in figure 1 with the 0 mapping with an arrow on the left to another 0, then mapping to a 1 with an arrow on the right. The bottom level, or leaves, from left to right is by definition the 1^{st} iteration of the TM sequence. We then construct the TM tree for the first iteration using the bitwise definition by starting with the T^0 TM tree which is simply 0.

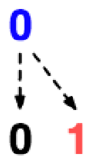

Figure 2: TM tree to the T^1 level via the bitwise definition

Then we find the bitwise negation for the entire tree, which is 1. We then place the T^0 tree on the left of its negation. The negation is shown in red in figure 2. We place a 0 above the 0 and 1 trees and connect the 0 to the 0 and 1 to complete the new composite tree, since $T^0 = 0$. We can read across the bottom levels of both trees (bitwise and substitution) and see that the resulting sequences are equivalent, which is due to the fact that the entire trees are equivalent. Next we will assume that this relationship holds for the $n = k$ iteration and show that, as a result, it holds for the $n = k + 1$ iteration. Given that we have the TM tree from the k^{th} iteration, we map each element of the bottom level using the homomorphism τ, which produces a new bottom level. By the substitution definition in section 2.4, the new bottom level of the TM tree from left to right is the $(k+1)^{st}$ iteration of the TM sequence, or T^{k+1}. The k^{th} row of the TM tree has the property that the second 2^{k-1} elements are the negation of the first 2^{k-1} elements. Therefore the $(k + 1)^{st}$ row, produced by the substitution using τ, will have the same property where the first 2^k elements are the negation of the second 2^k elements. This tree is equivalent to the $(k + 1)^{st}$ tree produced by the bitwise definition, where we take the TM tree from the k^{th} iteration and negate the entire tree. We place the k^{th} tree to the left of the negated tree, and place a 0 at the top to map to the 0 and 1 at the beginnings of the k^{th} tree and its negation respectively, as 0 is T^0.

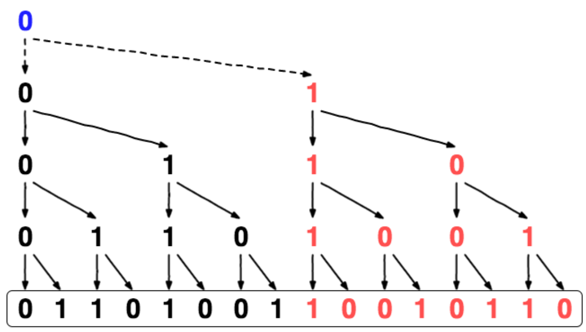

Figure 3: TM tree to the T^4 level via the substitution definition

This forms the $(k + 1)^{st}$ tree by the bitwise definition in section 2.1. The bottom level of this tree from left to right is by definition the $(k + 1)^{st}$ iteration of the TM sequence. The $(k+1)^{st}$ TM trees are equivalent and therefore by induction we have that the TM trees produced by both definitions at the k^{th} iteration are equivalent and therefore produce the same TM sequence.

□

3.2 Bitwise and Binary Sum Definitions

We will show that the bitwise negation definition in section 2.1 and the binary sum definition in section 2.3 are equivalent.

Proof. We will prove by induction that the bitwise and binary sum definitions are equivalent and therefore produce the same sequence. First note that the number of elements in the n^{th} iteration of the sequence is 2^n, meaning the number of elements doubles with every iteration. Also note that if we have the first 2^n integers, $\{0, 1, 2, ..., 2^n - 1\}$, we can generate the next 2^n integers by adding 2^n to each $i < 2^n$. This gives all integers $i < 2^{n+1}$, or the first 2^{n+1} integers, $\{0, 1, 2, ..., 2^{n+1} - 1\}$. Also note that all integers $i < 2^n$, in their binary representation, have a 0 in the place holder for 2^n, the n^{th} digit. For example $7 < 2^3 = 8$, the binary representation of 7 is 111, which we can also write as 0111 if we consider the 0 in the place holder for 2^3. This property is true for all integers $i < 2^n$ by the construction of the binary representation. Notice that adding 2^n to any integer $i < 2^n$ will then replace the 0 with a 1 in the 2^n place holder. For example $7 = 0111$ in binary and $7 + 2^3 = 7 + 8 = 15$, and $15 = 1111$ in binary.

The binary sum definition states that the sum (mod2) of the digits in the binary representation of the integer i is the i^{th} element of the TM sequence, i.e. $T_i = S_2(i) \bmod 2$ in equation 8. The binary representation of 0 is 0, which has the sum (mod2) of 0. Therefore $T_0 = S_2(0) \bmod 2 = 0$ by the binary sum definition. We also have that $T_1 = S_2(1) \bmod 2 = 1$. Then using equation 5 we have that the 1^{st} iteration of the sequence using the binary sum definition is $T_0, T_1 = 01 = T^1$. We also know that in the bitwise definition we begin with $T^0 = 0$, negate it to get $\overline{T^0} = 1$, which yields $T^1 = T^0\overline{T^0} = 01$, which is equivalent to T^1 from the binary sum definition.

We will assume that the sequences produced by both definitions are the same for the $n = k$ iteration, we will show that they are equivalent for the $n = k+1$ iteration. We start with the first 2^k elements of the TM sequence. These elements have been generated from the binary sum definition using the integers $i < 2^k$. To generate the second 2^k elements, we need the integers $2^k \leq i < 2^{k+1}$, or $\{2^k, 2^k + 1, ..., 2^{k+1} - 1\}$. As previously stated, we can do this by adding 2^k to each of the integers $i < 2^k$ which will add a single 1 digit to each of their binary representations. This negates the corresponding TM elements when we take the sum (mod2) of the digits of the binary representations. The first 2^k elements followed by the second 2^k elements make up the $n = k+1$ iteration of the TM sequence, and the second 2^k elements are the negation of the first 2^k elements using the binary sum definition. This is exactly the case with the bitwise negation definition which also yields the $n = k+1$ iteration of the TM sequence with the second 2^k elements which are the negation of the first 2^k elements. Therefore the bitwise negation definition and the binary sum definition are equivalent and yield the same sequence. □

3.3 Binary Sum and Recursive Definition

We will prove that the binary sum definition from section 2.3 and the recursive definition from section 2.2 are equivalent and therefore produce the same TM sequence.

Proof. From section 2.3 we have the properties that $S_2(2n) = S_2(n)$ and $S_2(2n+1) = S_2(n)+1 = S_2(2n)+1$. This corresponds directly to equations 6 and 7, which define the TM sequence in the recursive definition.

$$T_{2n} = S_2(2n) \bmod 2 \tag{14}$$
$$T_n = S_2(n) \bmod 2 \tag{15}$$
$$S_2(2n) \bmod 2 = S_2(n) \bmod 2 \tag{16}$$
$$\implies T_{2n} = T_n \tag{17}$$

This shows equivalence with equation 6. We can then show the equivalence with equation 7.

$$T_{2n+1} = S_2(2n+1) \bmod 2 = S_2(2n) + 1 \bmod 2 = S_2(n) + 1 \bmod 2 \tag{18}$$
$$T_n = S_2(n) \bmod 2 \tag{19}$$
$$\overline{T_n} = S_2(n) + 1 \bmod 2 \tag{20}$$
$$\implies T_{2n+1} = \overline{T_n} \tag{21}$$

We have $S_2(2n) + 1 \bmod 2$ is the negation of $S_2(n)$ which corresponds to the rule $T_{2n+1} = \overline{T_n}$. $S_2(2n) = S_2(n)$ is an equivalent way to write $T_{2n} = T_n$. □

4 The Prouhet-Tarry-Escott Problem

The Prouhet-Tarry-Escott, or PTE, problem is centered on a specific partitioning of integers. The underlying question is if it is possible to partition the integers $\{0, 1, 2, ..., 2^n - 1\}$ into two disjoint sets A and B such that

$$\sum_{j \in A} j^k = \sum_{j \in B} j^k \tag{22}$$

where j is an element of the set of integers $\{0, 1, 2, ..., 2^n - 1\}$. If such a partition exists, the next step is to determine for what values of k does equation 22 hold[1]. The trivial value of k, of course, is $k = 0$.

4.1 Partitioning of Numbers

To solve the Prouhet-Tarry-Escott problem, we will first consider the integers $0-7$, or $\{0, 1, ..., 2^3 - 1\}$. Can these integers be separated into two sets, A and B, such that the sum of the elements in the two sets are equal? Consider the partition given by the Thue-Morse sequence.

$$0, 1, 2, 3, 4, 5, 6, 7 \tag{23}$$

$$0, 1, 1, 0, 1, 0, 0, 1 \tag{24}$$

If the numbers $0-7$ are used as indices for the elements of the Thue-Morse sequence, then they can be partitioned where numbers that index a zero are in set A, and those that index a one are in set B. This partition yields $A = \{0, 3, 5, 6\}$ and $B = \{1, 2, 4, 7\}$.

$$0 + 3 + 5 + 6 = 14 \tag{25}$$

$$1 + 2 + 4 + 7 = 14 \tag{26}$$

We know it works for $k = 1$, and in the case above $k = 1$. This partition also works for $k = 2$

$$0^2 + 3^2 + 5^2 + 6^2 = 70 \tag{27}$$
$$1^2 + 2^2 + 4^2 + 7^2 = 70 \tag{28}$$

and $k = 3$

$$0^3 + 3^3 + 5^3 + 6^3 = 360 \tag{29}$$
$$1^3 + 2^3 + 4^3 + 7^3 = 416 \tag{30}$$

The sum of the elements in the two sets raised to the k power are equal for values $0 \leq k < 3$. The sums are not equal for $k = 3$, and notice that this was specifically for the partition of the numbers $\{0, 1, ..., 2^n - 1\}$ where $n = 3$. In general $0 \leq k < n$. Consider partitioning the set $\{x | x < 2^4, x \in \mathbb{N}\}$ using again the Thue-Morse sequence. Now $k = 3$ holds as $k = 3 < n = 4$.

$$0^3 + 3^3 + 5^3 + 6^3 + 9^3 + 10^3 + 12^3 + 15^3 = 7200 \tag{31}$$
$$1^3 + 2^3 + 4^3 + 7^3 + 8^3 + 11^3 + 13^3 + 14^3 = 7200 \tag{32}$$

Now consider from equation 22

$$\sum_{j \in A} j^k = \sum_{j \in B} j^k \tag{33}$$
$$\implies h^k \sum_{j \in A} j^k = h^k \sum_{j \in B} j^k \tag{34}$$
$$\implies \sum_{j \in A} h^k j^k = \sum_{j \in B} h^k j^k \tag{35}$$
$$\implies \sum_{j \in A} (h \cdot j)^k = \sum_{j \in B} (h \cdot j)^k \tag{36}$$

where $h \in \mathbb{R}$. This property shows that the numbers being partitioned do not need to be consecutive. They only need to be evenly spaced from one another. This is shown above as being evenly spaced by some number h. If $h \in \mathbb{Q}$, and therefore $h = \frac{q}{p}$ for $q \in \mathbb{N}$ and $p \in \mathbb{N}$, then we can deduce that $h \cdot j = r \implies r \in \mathbb{R}$.

Therefore, the numbers to be partitioned must be rational numbers separated by some constant h.

4.2 Polynomial Construction

Consider a general polynomial of the form

$$P(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_2 x^2 + a_1 x^1 + a_0 x^0 \tag{37}$$

If we assume the properties of the Thue-Morse partition expressed in equations 33 - 36 to be true then we can generalize the PTE problem to polynomials of degree no more than k. We will call this polynomial f. The polynomial can be constructed for values $0 \leq k < n$ as follows

$$\text{If } \sum_{k=0}^{n-1}\sum_{j\in A} a_k j^k = \sum_{k=0}^{n-1}\sum_{j\in B} a_k j^k \tag{38}$$

$$\text{then } \sum_{j\in A} f(j) = \sum_{j\in B} f(j) \tag{39}$$

where a_k is a constant coefficient that is specific to the k-power term of the polynomial constructed. Thus the Prouhet-Tarry-Escott partitioning problem has been generalized to a polynomial form.

4.3 Proving the PTE Problem

We will prove that the Thue-Morse sequence is a solution to the Prouhet-Tarry-Escott problem and satisfies the properties expressed in equations 22-39. We desire two disjoint sets A and B such that we get equality for equation 22. We can define sets X_k and Y_k as sets A and B respectively. Let X_k contain the indices $i < 2^{k+1}$ of the Thue-Morse sequence where $T_i = 0$ and let Y_k contain the indices $j < 2^{k+1}$ of the Thue-Morse sequence where $T_j = 1$. It is important to note that the Thue-Morse sequence has an equal number of 0s and 1s by the construction, and therefore X_k and Y_k have the same number of elements, i.e., $|X_k| = |Y_k|$[9].

Theorem 1. *For any polynomial f, of degree not exceeding k,*

$$\sum_{n\in X_k} f(n) = \sum_{n\in Y_k} f(n) \tag{40}$$

Proof. Honsberger[5] proves theorem 1 by induction on k. If $k = 0$, then equation 40 holds because $f(n)$ becomes a constant C and $|X_k| = |Y_k| = M$, and therefore we will have

$$C \cdot M = \sum_{n\in X_k} f(n) = \sum_{n\in Y_k} f(n) = C \cdot M \tag{41}$$

Then assume that equation 40 holds for a polynomial f of degree not exceeding k. We will then construct polynomials g and h where

$$g(n) = h(n + 2^{k+1}) - h(n) \tag{42}$$

and where h is a polynomial of degree no more than $k+1$. Therefore the difference cancels the highest power term, making g degree no more than k. We will assume

$$\sum_{n\in X_k} g(n) = \sum_{n\in Y_k} g(n) \tag{43}$$

for k and a polynomial g with a degree no more than k. For the induction proof we will need to show that under this assumption of k and g, it is implied that equation 40 also hold for a polynomial h with degree no more than $k+1$. It follows that

$$\sum_{n\in X_k} h(n + 2^{k+1}) - h(n) = \sum_{n\in Y_k} h(n + 2^{k+1}) - h(n) \tag{44}$$

$$\implies \sum_{n\in X_k} h(n + 2^{k+1}) - \sum_{n\in X_k} h(n) = \sum_{n\in Y_k} h(n + 2^{k+1}) - \sum_{n\in Y_k} h(n) \tag{45}$$

$$\implies \sum_{n\in X_k} h(n + 2^{k+1}) + \sum_{n\in Y_k} h(n) = \sum_{n\in Y_k} h(n + 2^{k+1}) + \sum_{n\in X_k} h(n) \tag{46}$$

Then using the property of the Thue-Morse sequence from equation 7 that if we shift a term in the first 2^{k+1} terms of the sequence by 2^{k+1}, the value is negated (0 for 1, 1 for 0). Then the equation becomes:

$$\implies \sum_{n\in(Y_{k+1}\setminus Y_k)} h(n) + \sum_{n\in Y_k} h(n) = \sum_{n\in(X_{k+1}\setminus X_k)} h(n) + \sum_{n\in X_k} h(n) \tag{47}$$

$$\implies \sum_{n\in Y_{k+1}} h(n) = \sum_{n\in X_{k+1}} h(n) \tag{48}$$

By induction the proof is complete as h is a general polynomial of degree no more than $k+1$ [5]. □

5 A Fair Sequence

In games which require participants to take turns, typically the advantage goes to the participant who has the first turn. This assumes that participants will alternate taking turns in an standard $ABABAB$... pattern. Is there a better way to sequence their turns to minimize the advantage of any specific participant? The solution is to take turns as dictated by the digits of the Thue-Morse sequence.

5.1 Pick-Up Basketball

Consider two team captains in a pick up basketball game who each wish to select a well-performing team. They will have to perform a draft to select their players. The pool of players from which they are selected have varying degrees of talent, which are known to the team captains. Naturally, the team captains will always choose the best remaining unselected player. The players are chosen in the order of highest talent to lowest talent. If we rank the players from worst to best, then the best players will have the largest rank number. For the sake of this example we will assume that the skill levels are also proportional to the rank number. If the team captains choose in the $ABABAB$... alternating pattern then their cumulative average skill are shown in Figure 4.

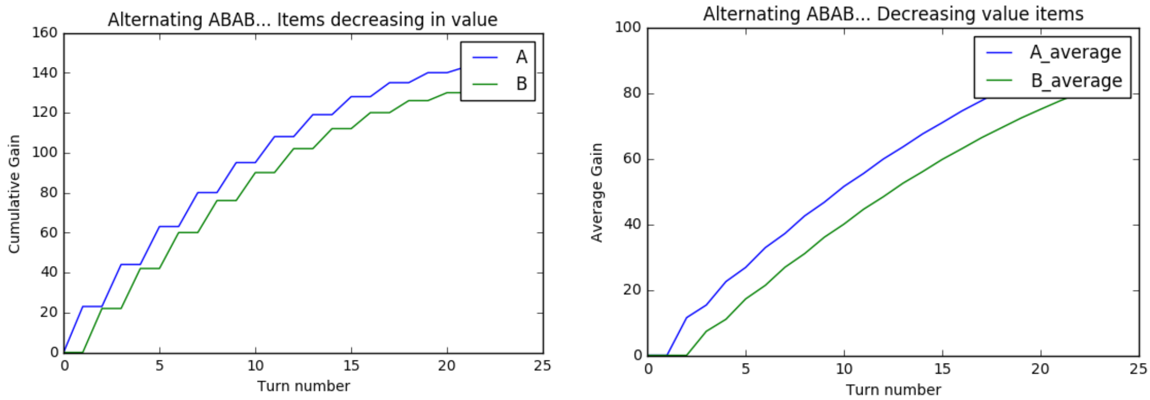

Figure 4: ABAB... Gain

It is clear that the captain that chooses first will always have the team with the higher skill level. The results of taking turns with the Thue-Morse sequence are show in figure 5

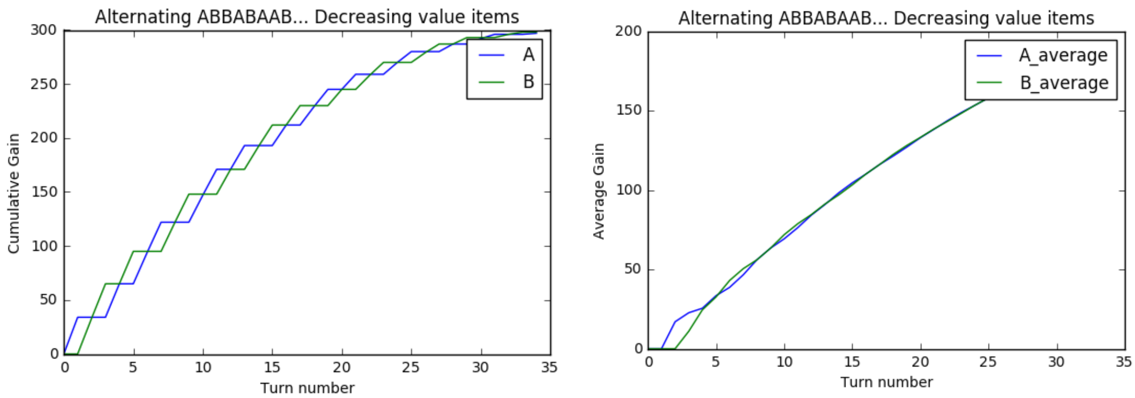

Figure 5: ABBA... (Thue-Morse) Gain

Using this pattern, the advantage of any one player starts to disappear. In fact, this is simply another version of the PTE partitioning problem where now the indices are values that are being evenly spread to participants turn by turn. This sequence is not the only fair sequence. We could consider the sequence where the captains still take turns based on the Thue-Morse sequence but instead of taking a single turn, they take m turns for every occurrence of a A or B. For example, where $m = 3$ we would have the sequence $AAABBBBBBAAA$... In that case, the sequence would converge to equality, but slower than would taking a single turn.

5.2 Pouring Two Cups of Coffee

We may also consider brewing a pot of coffee. When brewed, the first drops of coffee are the strongest, with every subsequent drop less potent than the last. This creates a gradient of coffee strength within the pot, with the strongest coffee at the bottom of the pot, and therefore the last to be poured out of the pot. Stirring the pot cannot eliminate the gradient strength. How can we pour two identical cups of coffee if there is a gradient of coffee strength within the pot? We can again discretize this problem, making it equivalent to the pick-up basketball problem, but then selecting the lowest number first, which parallels the weak coffee being poured first. Since these are equivalent problems, we see that alternating the coffee pours between the cups in an order dictated by the Thue-Morse sequence will yield two equivalent strength cups of coffee[2]. Richman [7] came to the same conclusion from taking the coffee pot problem one step further, making it a truly continuous gradient of strength. They show that as the number of iterations of the Thue-Morse sequence which are used increases, the difference in strength between the two cups of coffee tends toward 0.

6 Properties

6.1 Over-Lapping Property

Analysis of the Thue-Morse sequence has made a large contribution to the study of cube-free and square-free words in combinatorics. Norwegian mathematician Axel Thue had originally constructed the sequence because it satisfied the overlap-free property[4][6]. We say that a sequence (or string) over an alphabet is *square-free* if no two non-empty adjoining sub-strings are equal and *cube-free* if no three non-empty adjoining sub-strings are equal[6]. If a word is *overlap-free* then it contains no sub-block of the form $awawa$, where $a \in \{0, 1\}$ and w is a binary block[1].

Theorem 2. ***Thue***: *The Thue-Morse sequence is overlap-free.*

The Thue-Morse sequence is a cube-free binary sequence which can be used to derive a square-free sequence. We will define a new sequence v. For $n \geq 1$ let v_n be explicitly defined as the number of 1's between the n^{th} and the $(n + 1)^{st}$ 0 in the

Thue-Morse sequence. We get

$$v = 21020121012... \tag{49}$$

Thue then proved that the sequence v is square-free. This finding launched the field of *combinatorics on words*[1].

6.2 Application to Chess

There is a rule in chess which states that a draw occurs if the same sequence of moves occurs three times in a row. If one can avoid this rule, then there is the possibility for an infinite game of chess. Dutch chess grandmaster Machgielis Euwe derived a cube-binary pattern to try to avoid a draw. He mapped a 0 in the patturn to one set of chess moves, and a 1 in the pattern to a difference set of moves, thus ensuring that the game would never come to a draw due to the cube-free property of the sequence he derived. The sequence he derived was coincidentally exactly the Thue-Morse sequence[1].

6.3 Alternating Symmetry

The Thue-Morse sequence also exhibits alternating symmetry. Consider what the sequence looks like at each intermediate iteration during its construction. The iterations are separated below starting with the $n = 0$ iteration and ending with the $n = 6$ iteration.

$$0$$

$$0|1$$

$$01|10$$

$$0110|1001$$

$$01101001|10010110$$

$$0110100110010110|1001011001101001$$

$$01101001100101101001011001101001|10010110011010010110100110010110$$

If we then consider then folding the sequence at the middle where the bars are placed, and letting the two halves overlap, an alternating pattern emerges. Whenever $n \bmod 2 = 0$, the 0's pair with 0's and the 1's pair with 1's, and thus summing the overlapping digits mod 2 yields all 0's. Then whenever $n \bmod 2 = 1$, the 0's pair with 1's and the 1's pair with 0's, and thus summing the overlapping digits mod 2 yields all 1's. There is mirror symmetry between the two halves at every iteration, with the odd iterations having their symmetry with the bitwise negation of the other half.

6.4 Generating Function

The Thue-Morse sequence elements can be used as coefficients to construct a power series over the finite field $\mathbb{GF}(2)$.

$$F(x) = 0 + 1x + 1x^2 + 0x^3 + 1x^4 + 0x^5 + ... \tag{50}$$

The function $F(x)$ is one of two solutions to the quadratic equation

$$(1+x)F^2 + F = \frac{x}{1+x^2} \bmod 2 \tag{51}$$

the other solution is $\overline{F}$, the conjugate of F, where all 0 coefficients are replaced with 1s and vice versa. We will then refer to the power series F(x) and other power series by their coefficients.

$$F = (0110100110010110...) \tag{52}$$

First we need to get F^2 which requires squaring the series. The rule for squaring over $\mathbb{GF}(2)$ is

$$(A+B)^2 = (A^2 + B^2) \bmod 2 \tag{53}$$

When applied as F^2, this has the affect of inserting 0's at the odd indices yielding

$$F^2 = (0010100010000010100...) \tag{54}$$

Then multiplying by x simply adds a 0 to the front.

$$xF^2 = (00010100010000010100...) \tag{55}$$

Then adding F^2 gives

$$(1+x)F^2 = (0011110011000011...) \tag{56}$$

The next step is to add $F(x)$ which gives

$$(1+x)F^2 + F = (0101010101010101010101...) \tag{57}$$

We know the expansion of the geometric power series

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + x^4 + ... \tag{58}$$

we can therefore derive

$$\frac{1}{1+x^2} = 1 + x^2 + x^4 + x^6 + ... \tag{59}$$

$$\implies \frac{x}{1+x^2} = x + x^3 + x^5 + x^7 + ... \tag{60}$$

Reverting back to the coefficients form of the polynomial from equation 57 and setting equal to equation 60 shows equality of equation 51[8].

$$(1+x)F^2 + F = (0101010101010101010101...) = 0 + 1x + 0x^2 + 1x^3 + 0x^4 + 1x^5 + ... \tag{61}$$

which shows

$$(1+x)F^2 + F = \frac{x}{1+x^2} \bmod 2 \tag{62}$$

7 Extending The Sequence

All of the work to this point has been done with single dimensional sequences in base 2. That is the form of the original Thue-Morse sequence, however many of the rules that define the sequence can be generalized to other bases and dimensions.

7.1 Generalizing To Other Bases

A bitwise negation does not make sense outside of base 2, but the rule can be slightly modified to make the bitwise definition more robust and compatible in base m. Instead of a bitwise negation, a digit shift $(\mathrm{mod} m)$ will be used to create multiple shifts of a sequence at a given iteration. That means if we are working in base m then the operation $+i(\bmod m)$ will be performed on all elements of the sequence T^n for $0 \leq i < m$, creating m new shifted sequences, $\overline{T^n}^i$. All of these shifted sequences are then concatenated together.

$$T^n = \overline{T^{n-1}}^0 \overline{T^{n-1}}^1 ... \overline{T^{n-1}}^{m-2} \overline{T^{n-1}}^{m-1} \tag{63}$$

where T^n is partial Thue-Morse sequence at the the n^{th} iteration of the rule. In base two this operation is equivalent to doing the bitwise negation from section 2.1. rewriting the definition to this more general form allows for a sequence outside of base 2. For example for base $m = 3$. Let $T^0 = 0$.

$$\overline{T^0}^0 = 0 + 0 = 0 \bmod 3 \tag{64}$$

$$\overline{T^0}^1 = 0 + 1 = 1 \bmod 3 \tag{65}$$

$$\overline{T^0}^2 = 0 + 2 = 2 \bmod 3 \tag{66}$$

$$\implies T^1 = \overline{T^0}^0 \overline{T^0}^1 \overline{T^0}^2 = 012 \tag{67}$$

$$\overline{T^1}^0 = 012 \bmod 3 \tag{68}$$

$$\overline{T^1}^1 = 120 \bmod 3 \tag{69}$$

$$\overline{T^1}^2 = 201 \bmod 3 \tag{70}$$

$$\implies T^2 = \overline{T^1}^0 \overline{T^1}^1 \overline{T^1}^2 = 012120201 \tag{71}$$

This can also be done using a substitution. The general substitution rule can be defined by a morphism $\tau : \{0, 1, ..., n-1\} \rightarrow \{0, 1, ..., n-1\}$. For the base $m = 3$, let τ be defined by $0 \mapsto 012$, $1 \mapsto 120$, and $2 \mapsto 201$ [4]. Then, iterating the morphism

τ on the initial symbol 0 gives

$$0 \mapsto 012 \mapsto 012120201 \mapsto 012120201120201012201012120 \tag{72}$$

Now we have a way to evenly pour coffee into m cups.

7.2 Generalizing To Other Dimensions

The sequence can be generalized to d-dimensions. First extend d-perpendicular axes from a point. Then form a single dimensional Thue-Morse sequence in base-m. We can call this sequence T_m where the i^{th} element of the sequence is $T_{m,i}$. Then let $\vec{v} = (x_0, x_1, ..., x_{d-1})$ represent a coordinate in the d-dimension space. We will then define a function $TM(\vec{v})$, $\mathbb{Z}^d \to \mathbb{Z}_m$ which gives the Thue-Morse value for a given point defined by a vector $\vec{v}$.

$$TM(\vec{v}) = \left(\sum_{x_i \in \vec{v}} T_{m,x_i} \right) \bmod m \tag{73}$$

It is hard to visualize beyond 2 dimensions. Figures 6 and 7 show that some interesting patterns emerge even in 2 dimensions.

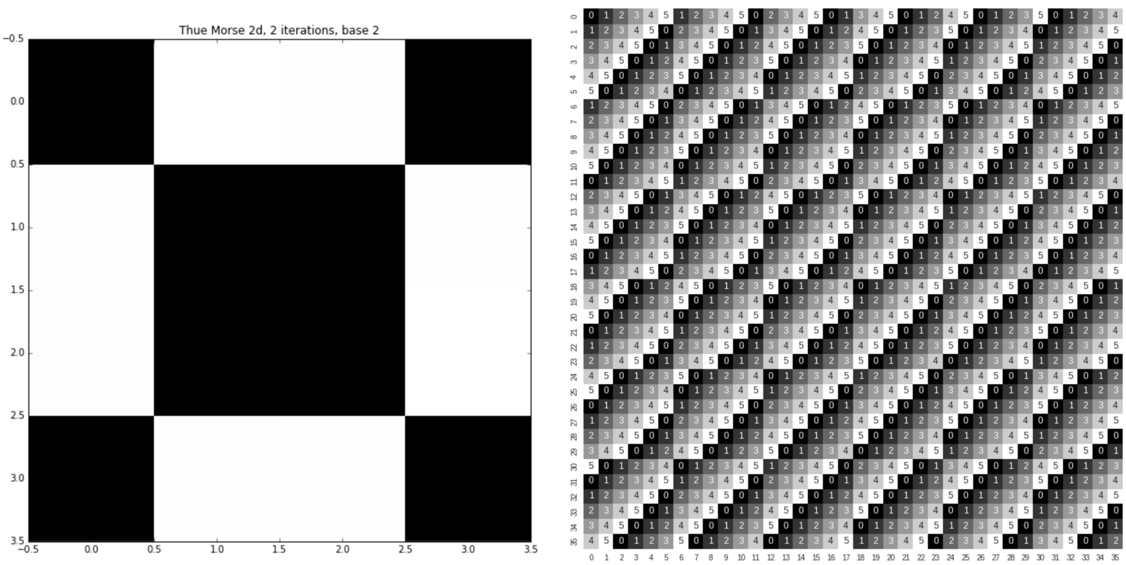

Figure 6: 2 Iterations in Base 2 and 6

Figure 6 shows a black and white illustration of the 2 dimensional, 2 iteration Thue-Morse sequence in base 2 on the left. The black squares represent 0s and the white squares represent 1s. The full image corresponds to the numbers in the table below.

0	1	1	0
1	0	0	1
1	0	0	1
0	1	1	0

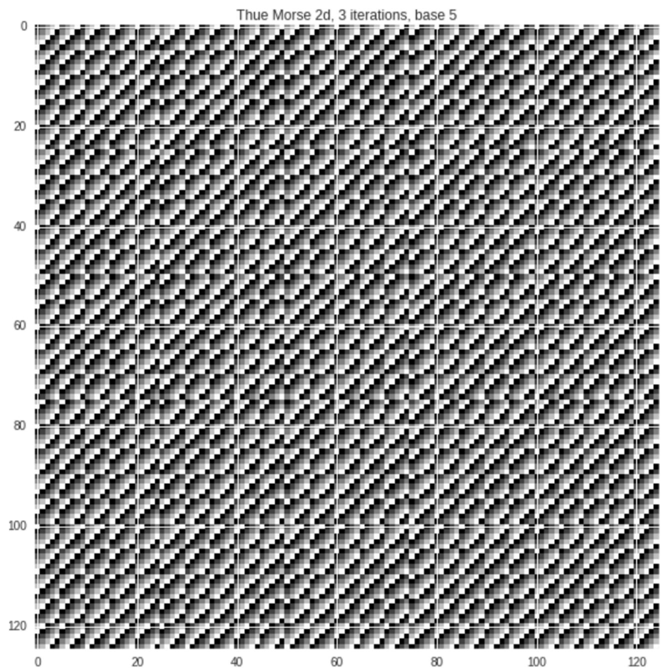

Figure 7: 3 Iterations in Base 5

8 Conclusion and Reflection by the Author

The Thue-Morse sequence is a fascinating sequence that does not choose any favorites. It can pour equal strength cups of coffee, pick basketball teams, or prevent you from losing a game of chess. One of the most critical aspects of this topic is the recursive definition. It is what the Thue-Morse sequence is build upon. I have studied recursive and explicit definitions when initially learning about sequences and series in Calculus II. Then I saw these types of definitions again when working on my Collatz Conjecture project. The sequences produced by the rules used in the Collatz Conjecture can only be defined recursively, which is a contrast to the Thue-Morse sequence. It can be defined recursively or explicitly, similar to the Fibonacci sequence which I learned when taking abstract algebra.

Induction is another critical tool that analysis of the Thue-Morse sequence relies on. At the beginning of my math career I was not very familiar with induction let alone proof writing. I would not have been able to explore such a topic had I not studied logic and proof writing in discrete mathematics and abstract algebra. Success in these courses hinges on a strong understanding of mathematical concepts from algebra and calculus. Some portions of this project briefly venture into number theory and combinatorics, again both topics that I was first exposed to through discrete mathematics. In discrete mathematics I also focused a lot on set theory. The Prouhet-Tarry-Escott problem is mainly a question of number theory and set theory. An understanding of sets and partitions is necessary for working with the PTE problem.

I believe that the most useful field of knowledge I drew upon during this study was abstract algebra. Almost every concept I looked at had at least a touch of abstract algebra. The recursive definitions, mappings, morphisms, generating functions, and finite fields are all abstract algebra topics. When I took abstract algebra, I took an applied course. I think that some of the things I saw in the course were very specific and practical, like coding theory, encryption, permutations, and other topics. I also studied modular arithmetic for a while in the course. Modular arithmetic is crucial to many of the definitions used for the Thue-Morse sequence. I think that this study of the Thue-Morse sequence helped me fully branch the gap from the applied to the theoretical abstract algebra and now I have a much stronger knowledge of what it's all about. There are also ideas from multivariable calculus and linear algebra hidden throughout the work. The multidimensional sequences can be represented using vectors and matrices, with explicit definitions for each entry in the matrix. Section 7 utilizes vectors and multivariable concepts to take the specifics of the sequence into a general form.

My course work at Ithaca college taught me how to work through very interesting yet challenging problems such as this one. I can already see the difficult undergraduate work paying off in my senior course work, research work, and even in job interviews with difficult brain teasers. I also give a special thanks to professor Ted Galanthay for guiding me through the work on this topic.

References

[1] J.-P. Allouche and J. Shallit, *Sequences and Their Applications: The Ubiquitous Prouhet-Thue-Morse Sequence*, Springer, 1999.

[2] E. D. Bolker, C. Offner, R. Richman, and C. Zara, *The prouhet-tarry-escott problem and generalized thue-morse sequences*, Journal of Combinatorics, 7 (2016), pp. 117–133.

[3] S. Brlek, *Enumeration of factors in the thue-morse word*, Discrete Applied Mathematics, 24 (1989), pp. 83–96.

[4] M. Bucci, N. Hindman, S. Puzynin, and L. Q. Zamboni, *On additive properties of sets defined by the thue-morse word*, Journal of Combinatorics Theory, 120 (2013), pp. 1235–1245.

[5] R. Honsberger, *Mathematical Diamonds*, Mathematical Association of America, 2003.

[6] C. D. Offner, *Repetitions of words and the thue-morse sequence*, (2007).

[7] R. M. Richman, *Recursive binary sequences of differences*, Complex Systems, 13 (2001), pp. 381–392.

[8] E. W. Weisstein, *Thue-morse sequence*. http://mathworld.wolfram.com/Thue-MorseSequence.html. Accessed: 2016-12-16.

[9] C. Williamson, *An overview of the thue-morse sequence*, (2012).