An Overview of the Thue-Morse Sequence

Christopher Williamson

June 4, 2012

Contents

1 Introduction

The Prouhet-Thue-Morse sequence, more commonly called the Thue-Morse (TM) sequence, is defined on the alphabet $\sum = \{0, 1\}$ and was first considered by Prouhet in 1851 [9]. The sequence has applications in numerous fields of mathematics, and despite the simplicity of the terms in the sequence, we will see that its properties are anything but trivial.

2 Equivalent Definitions

There are many different ways to define the TM sequence. Of course, they are all equivalent, but this is not always obvious. In this section, we will review the most commonly seen definitions.

2.1 Sum of Terms Definition

Definition 1. *If T_j is the j^{th} number in the TM sequence, then T_j equals the number of ones that occur in the base 2 expansion of j, mod 2.*

For example, $T_{11} = 1$, as $(11)_{10} = (1011)_2$ and 1011 has three ones, which is congruent to one (mod two).

Frequently, the terms of the sequence are seen concatenated together without commas. Here is the beginning of the sequence, written in that form. The reason for the concatenation will become apparent shortly.

011010011001011010010110...

A pair of observations leads to another definition of this sequence. Note that $T_n = T_{2n}$, as the multiplication by two simply appends a zero to the binary number, leaving the number of ones unchanged. (Obviously, this statement implies that $T_n = T_{2^j n}$ for any integer j greater than zero.) Similarly, $T_n \neq T_{2n+1}$, and since the sequence consists of only two characters, this is enough to determine the value of T_{2n+1} from the value T_n. It is easy to verify that the two above observations are strong enough to define a sequence of their own, even if we had no previous knowledge of the TM sequence (although it is necessary to define a starting value and say that the alphabet size is two).

2.2 Recursive Definition

Definition 2. *The sequence defined on the alphabet $\{0,1\}$ by the rules: $T_0 = 0$, $T_n = T_{2n}$, and $T_n \neq T_{2n+1}$ is the Thue-Morse sequence.*

With definition 2 established, another construction of the sequence becomes fairly obvious. Say that the first 2^n terms of the sequence have been written down, and express this partial sequence as TM_{2^n}. Then construct the next 2^n terms by writing the same sequence, TM_{2^n} again, except with the numbers flipped. This new sequence of 2^n numbers, which is then appended to the end of the previously found terms, is represented as $\overline{TM_{2^n}}$, so that the sequence after one step has become the concatenation of the sequences: $TM_{2^n}\overline{TM_{2^n}}$. One more step would produce: $TM_{2^n}\overline{TM_{2^n}}\,\overline{TM_{2^n}\overline{TM_{2^n}}}$, which is equivalent to $TM_{2^{n+1}}\overline{TM_{2^{n+1}}}$, which in turn equals $TM_{2^{n+2}}$. Steps 1-4 of this algorithm are:

0

01

0110

01101001

0110100110010110

Note that in step 1, we transform the sequence from 0 to 01. It is clear that in step j, we go from a sequence of length 2^{j-1} to a sequence of length 2^j, by taking $T_n = \overline{T_{n+2^j}}$. This equality does not hold in general, but is true during step j. The reason for this is that the largest n value that could be present during step j is 2^{j-1}. Then, by adding 2^j, we are simply adding a one to the left of the number, which flips the value of the sequence. This flipping algorithm leads to the third definition.

2.3 Flipping Definition

Definition 3. *Let $TM_0 = 0$ and $TM_{2^{n+1}} = TM_{2^n}\overline{TM_{2^n}}$. Then, the Thue-Morse sequence equals TM_∞.*

This is why the Thue-Morse sequence is sometimes called the Thue-Morse infinite word, and why the commas are removed in favor of concatenation. One can define a language as all the subsequences (starting at 0) of the TM sequence, meaning that the words for the language are TM_n for every nonnegative integer n. The final definition to be mentioned is as follows.

2.4 Morphism Definition

Definition 4. *Let μ be a sequence substitution map defined as $\mu(0) = 01$ and $\mu(1) = 10$. Then, the Thue-Morse sequence can be defined as $\mu^\infty(0) = \mu(\mu(\mu(...\mu(0)...)))$, which is the substitution map applied infinitely many times to the starting sequence 0.*

Note that this means that one can write a finite sequence of zeroes and ones that converges to the Thue-Morse infinite word by writing down zero, and applying the map forever, an action which consists of repeatedly replacing zeroes with 01 and ones with 10. At first glance, it may not even be clear that the sequence limit $\mu^\infty(0)$ is well defined in that it converges. Convergence in this context means that for a given k, there exists an ω such that the first k terms of the sequence are the same for all morphisms of the sequence after ω (meaning that the first k terms are the same for $\mu^\omega(0), \mu^{\omega+1}(0), \mu^{\omega+2}(0), \mu^{\omega+3}(0), ...$). However, it is readily apparent that the first term of $\mu^\omega(0)$ for any ω is 0, and it isn't too hard to see that if the n^{th} term of the sequence converges for all ω after a certain point, that the $(n+1)^{th}$ and $(n+2)^{nd}$ terms must converge also. This forms the basis for an inductive proof that $\mu^\infty(0)$ converges. We will now see that from this definition, one can derive the relations seen in definition 2, and conclude that this definition is consistent with the other three.

Since $\mu^\infty(0)$ converges to some sequence, it is legal to say in some sense that $\mu^\infty(0) = \mu^{\infty+1}(0)$. This simply represents the fact that no terms between the sequences can differ, as this would mean that there existed a k such that T_k did not hold the same value for all $\mu^\omega(0)$ for all ω greater than some integer, a contradiction. We can take advantage of this fact − since applying the morphism to a given sequence simply doubles the index of a given term in the original sequence, we arrive at the fact that $T_n = T_{2n}$. Also, when one applies the morphism to $\mu^\omega(0)$, the $(2n+1)^{th}$ term of $\mu^{\omega+1}(0)$ coincides with the second number in the morphism mapping of the n^{th} term of $\mu^\omega(0)$, which is always the opposite number of the n^{th} term of $\mu^\omega(0)$. Thus, $T_n \neq T_{2n+1}$, as we know that the sequence of repeated substitutions converges. Then, since the first term is always 0, we see that this definition is equivalent to definition 2. There is one interesting consequence of this definition–that the Thue-Morse sequence is a fractal sequence.

Definition 4.1 *A sequence is fractal if it displays self similarity, in that there exists a modification that can be made to the sequence that results in the same sequence again.*

Theorem 1. *The Thue-Morse sequence is a fractal sequence, with the rule that the removal of* *every odd indexed term results in the original sequence.*

Proof. The removal of all the odd-indexed terms from the Thue-Morse sequence coincides exactly with the transformation from $\mu^{\infty}(0)$ to $\mu^{\infty-1}(0)$. These two are the same sequence, by the convergence of the sequence of sequences $\mu^{\omega}(0)$. Alternatively, recall that $T_n = T_{2n}$. Then, the removal of the terms at odd indices causes the remaining terms' indices to halve, which does not change the sequence.

2.5 Infinite Product Definition

First, define a sequence with terms a_j (which equal 0 or 1) such that the following equation holds for values of x that yield convergence:

$$\prod_{i=0}^{\infty}(1-x^{2^i}) = \sum_{j=0}^{\infty}(-1)^{a_j}x^j$$

By considering expanding out the terms,

$$(1-x)(1-x^2)(1-x^4)(1-x^8)...$$

one sees that the sum will contain monomials of every possible positive integer power, as x^n is formed by multiplying out all the terms with an exponent that occurs in the binary expansion of n.

For example, if $n = 9$, then the monomial x^n is formed by multiplying together x and x^8, as the base-2 expansion of 9 contains ones in precisely the 1's place and the 8's place. Since every number has a unique expression in base 2, and since every power of 2 is available in the infinite product, a term x^n exists for all positive integers n (and it has a unique coefficient).

Now, we must consider the coefficients of these terms. Since the formation of x^n involves the multiplication of a number of terms equal to the number of ones in its binary expansion, we arrive at the fact that the coefficient is simply equal to $(-1)^Y$, where Y is the number of ones in the binary expansion. So, if a number n has an even number of ones in its base-2 expansion, $a_n = 0$ and an odd number of ones yields $a_n = 1$. Note that by definition 1, $T_n = a_n$. We now arrive at the fifth definition of the Thue-Morse sequence.

Definition 5. *The terms of the Thue-Morse sequence, T_j, defined with values {0,1} are the numbers such that the following equation holds for values of x that allow the product and sum to converge.*

$$\prod_{i=0}^{\infty}(1-x^{2^i}) = \sum_{j=0}^{\infty}(-1)^{T_j}x^j$$

3 The Prouhet-Tarry-Escott Problem

The PTE problem, also commonly called the multigrades problem, asks for two disjoint sets A and B of n integers such that for all integers k in a certain range,

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

The Thue-Morse sequence is a valuable tool for solving a very general version of this problem.

3.1 Application of the TM sequence

The TM sequence can be used to solve this problem by defining the sets A and B to be sets of indices depending on the value of the sequence. Due to the infinite length of the TM sequence, these sets can becomes arbitrarily large, and if both sets contains 2^c integers, then we will be able to have k range from 0 to c. We will see that application of the TM sequence requires that the sets contain a power of two elements, but we will be able to generalize from simply taking powers of the elements to evaluating general polynomials of degree from 0 to c at the elements.

First, define the sets X_k and Y_k (taking the names of A and B). Let X_k contain the indices i less than 2^{k+1} where T_i is zero and Y_k contain the indices j less than 2^{k+1} where T_j is 1. Note that in the notation of definition 3, TM_1 has an equal number of zeroes and ones among its 2^1 terms. Since each subsequent TM_k is formed by concatenating TM_{k-1} and the reverse of TM_{k-1}, each TM_k has an equal number of zeroes and ones. The implication of this is that among the first 2^{k+1} terms of the Thue-Morse sequence (for $k \geq 0$), there are 2^k zeroes and the same number of ones. So, we have: $|X_k| = |Y_k|$. Then, Honsberger [6] proves the following.

Theorem 2. For any polynomial f of degree not exceeding k,

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

Proof. The proof is straightforward, and by induction on k. First, note that if $k = 0$, the property holds because there are the same number of terms in X_k and Y_k. Now, assume the truth of the property for polynomials of degree not exceeding k, and let g be a polynomial that gives the following difference:

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

where h is an arbitrary polynomial of degree at most $k + 1$. Then, due to the cancellation of the highest power terms, g is of degree at most k, and we can assume that:

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

It follows that

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

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

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

It follows from definition 3 that if one shifts a term from the first 2^{k+1} terms in the TM sequence by 2^{k+1}, that the value flips. Therefore, the above equation becomes

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

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

The theorem is then proved, as h is a general polynomial of degree at most $k + 1$.

Example.

Let $f(x) = x^2 + x + 2$ (note that it is not necessary for the coefficients to be integers). Take k to be 2, which means that $X_2 = \{0, 3, 5, 6\}$ and $Y_2 = \{1, 2, 4, 7\}$. Then, we have:

$$f(0) + f(3) + f(5) + f(6) = 2 + 14 + 32 + 44 = 92 = 4 + 8 + 22 + 58 = f(1) + f(2) + f(4) + f(7)$$

4 An Interesting Integer Partition

If one recalls the definition given in the previous section of X_k and Y_k, then one can consider X_∞ and Y_∞ (to be renamed X and Y). These are the infinite sets that contain the indices at which the TM sequence is zero (in the case of X) or one (in the case of Y). Another interesting problem that can be solved via the TM sequence is to create a partition of the non-negative integers such that given an integer n, n can be represented as a sum of two distinct numbers in the same number of ways for each set. Call this the property of an integer partition property P. Then, we have another theorem.

Theorem 3. A partition of the non-negative integers has property P if and only if that partition is X, Y (proved by Lambek and Moser in 1959 [7]).

4.1 Solution Existence

Let X and Y be our two partitions, defined as described above. Also, let the countable number of elements of X (resp. Y) be denoted as x_j (resp. y_j). Then, define two functions, $X(t)$ and $Y(t)$ for t values that yield convergence, where

$$X(t) = \sum_{x_j \in X} t^{x_j} = t^0 + t^3 + t^5 + t^6 + ...$$

$$Y(t) = \sum_{y_j \in Y} t^{y_j} = t^1 + t^2 + t^4 + t^7 + ...$$

Note that the section justifying definition 5 allows us to say that when $\prod_{i=0}^{\infty}(1 - t^{2^i})$ is expanded out into a sum, the terms with negative coefficient (which must be -1) correspond to those with an odd number of ones in the binary expansion of the exponent. Likewise, the terms with a positive coefficient (namely, 1), have an exponent with an even number of ones in the binary expansion. Since X is the set with the numbers that have an even number of ones and Y is the set of numbers with an odd number of ones, we arrive at:

$$\prod_{i=0}^{\infty}(1 - t^{2^i}) = X(t) - Y(t).$$

Recalling the sum of the geometric series, and noting that every non-negative integer is represented once as an exponent of a term in either $X(t)$ or $Y(t)$, we also know that

$$X(t) + Y(t) = \frac{1}{1 - t}$$

We can combine these facts to determine that

$$(X(t) + Y(t))(X(t) - Y(t)) = \prod_{i=1}^{\infty}(1 - t^{2^i}) = (1 - t^2)(1 - t^4)(1 - t^8)...$$

Note that we can expand out the terms of the left-most term in the set of above equalities to obtain $X^2(t) - Y^2(t)$. Also, by considering

$$\prod_{i=0}^{\infty}(1 - t^{2^i}) = X(t) - Y(t).$$

but replacing t with t^2, we see that

$$\prod_{i=1}^{\infty}(1-t^{2^i}) = X(t^2) - Y(t^2).$$

Finally, it is now apparent that

$$X^2(t) - Y^2(t) = X(t^2) - Y(t^2) \Longrightarrow X^2(t) - X(t^2) = Y^2(t) - Y(t^2).$$

Considering the functions

$$X^2(t) = (t^0 + t^3 + t^5 + t^6 + ...)(t^0 + t^3 + t^5 + t^6 + ...)$$

and

$$Y^2(t) = (t^1 + t^2 + t^4 + t^7 + ...)(t^1 + t^2 + t^4 + t^7 + ...),$$

we can see that the coefficient on some monomial t^m for odd m in the first (or second) expansion is simply twice the number of distinct ways that two exponents in the equations (or elements in the set X or Y) can be added to make m. For example, t^8 will have a coefficient of 2 in $X^2(t)$ due to the following two multiplications: t^3t^5 and t^5t^3. If m is even, then for a given $X^2(t)$ or $Y^2(t)$, nothing is different unless $\frac{m}{2}$ occurs as an exponent in whichever function we are considering (before expansion). For example, this occurs for $m = 6$ in $X^2(t)$, as a t^3 term is present before any expansion. In this case, the coefficient for t^m will be one greater than is desired, as we want the coefficient to give the number of ways of adding two *distinct* numbers in X or Y to reach m. So, the subtraction of the function $X(t^2)$ from $X^2(t)$ and of the function $Y(t^2)$ from $Y^2(t)$ corrects the coefficients by subtracting one from the coefficients of terms with exponents m where $\frac{m}{2}$ also occurs in the same set as m. Since the equality of

$$X^2(t) - X(t^2) = Y^2(t) - Y(t^2)$$

has been established, we know that these coefficients (which give twice the number of ways that m can be represented) are the same for both sets, meaning that X and Y give a partition of the integers that has the desired property – any number m can be represented as the sum of two distinct members from X or from Y in the same number of ways.

4.2 Solution Uniqueness

We will prove the uniqueness of this partition just as we usually do – assume that there is another partition with the same property and show that it must be equal to the previous partition X, Y.

Let C and D be a new partition of the non-negative integers, and without loss of generality, assume that $0 \in C$. We now define functions related to the sets in the same ways before, except now we do not know exactly what numbers are in the exponents.

$$C(t) = \sum_{c_j \in C} t^{c_j} = t^{c_1} + t^{c_2} + t^{c_3} + t^{c_4} + ...$$

$$D(t) = \sum_{d_j \in D} t^{d_j} = t^{d_1} + t^{d_2} + t^{d_3} + t^{d_4} + ...$$

Since C and D are a partition of the integers, we again have

$$C(t) + D(t) = \frac{1}{1-t}.$$

Using the exact same reasoning as before, we know that the coefficients on a given monomial in the expansion of $C^2(t) - C(t^2)$ or $D^2(t) - D(t^2)$ give twice the number of distinct ways of summing two different numbers in C or D to obtain the exponent. Since we have assumed that C and D give another partition with the desired property, we can state the following as an equality, and continue working to find that the C, D partition is the same as the X, Y partition.

$$C^2(t) - C(t^2) = D^2(t) - D(t^2)$$

By rearranging this so that the terms with the function squared are on the same side and the terms with the argument squared are on the opposite side, and then factoring, we reach

$$(C(t) + D(t))(C(t) - D(t)) = C(t^2) - D(t^2),$$

which implies

$$C(t) - D(t) = (1 - t)(C(t^2) - D(t^2)).$$

Note that t can be replaced with t^2 or t^4 or t^{2^j} to obtain

$$\begin{gathered}
C(t^2) - D(t^2) = (1 - t^2)(C(t^4) - D(t^4)), \\
C(t^4) - D(t^4) = (1 - t^4)(C(t^8) - D(t^8)), \\
\vdots \\
C(t^{2^j}) - D(t^{2^j}) = (1 - t^{2^j})(C(t^{2^{j+1}}) - D(t^{2^{j+1}})).
\end{gathered}$$

We can see that by starting with the equation for $C(t) - D(t)$ and by replacing the right most term, $C(t^2) - D(t^2)$, with the expression $(1 - t^2)(C(t^4) - D(t^4))$ we obtain

$$C(t) - D(t) = (1 - t)(1 - t^2)(C(t^4) - D(t^4)).$$

This process can be continued indefinitely, by continually replacing the right-most term of the current equation by the right side of the next equation. Thus,

$$C(t) - D(t) = \left(\prod_{j=0}^{N}(1 - t^{2^j})\right)(C(t^{2^{N+1}}) - D(t^{2^{N+1}}))$$

for any arbitrarily large N. Note that we can take the limit as N tends to infinity, to obtain:

$$C(t) - D(t) = \prod_{j=0}^{\infty}(1 - t^{2^j}).$$

More formally, the absolute value of the difference δ_N between $C(t) - D(t)$ and the partial product up to N is:

$$\delta_N = \left|[C(t) - D(t)] - \prod_{j=0}^{N}(1 - t^{2^j})\right| = \left|(C(t^{2^{N+1}}) - D(t^{2^{N+1}}))\prod_{j=0}^{N}(1 - t^{2^j}) - \prod_{j=0}^{N}(1 - t^{2^j})\right|$$

$$\delta_N = \left|\prod_{j=0}^{N}(1 - t^{2^j})\Big((C(t^{2^{N+1}}) - D(t^{2^{N+1}})) - 1\Big)\right| = \left|\prod_{j=0}^{N}(1 - t^{2^j})\right|\left|\Big((C(t^{2^{N+1}}) - D(t^{2^{N+1}})) - 1\Big)\right|$$

Since the t values were chosen so that the infinite product converges, we can say that there exists a number M such that $M \geq \left|\prod_{j=0}^{N}(1 - t^{2^j})\right|$ for all N. Thus,

$$\delta_N \leq M\left|\Big((C(t^{2^{N+1}}) - D(t^{2^{N+1}})) - 1\Big)\right|$$

Note that $C(t)$ and $D(t)$ also are only defined for t values where the functions converge uniformly, perhaps by requiring that $t \in (0, 1-\epsilon]$ for a positive ϵ with absolute value less than one. Since

$$\lim_{N\to\infty} (C(t^{2^{N+1}}) - D(t^{2^{N+1}})) = 1$$

for $t \in (0, 1-\epsilon]$ (as every term vanishes except for the t^0 term in $C(t)$), we have:

$$\lim_{N\to\infty} \delta_N \leq \lim_{N\to\infty} M((C(t^{2^{N+1}}) - D(t^{2^{N+1}})) - 1) = 0$$

So, it is legal to say that

$$C(t) - D(t) = \prod_{j=0}^{\infty} (1 - t^{2^j}).$$

The terms t^m in the expansion of the above product with positive (resp. negative) coefficients have exponent $m \in C$ (resp. in D) and recalling definition 5, we know that the terms t^m with positive (negative) coefficient have an even (odd) number of ones in the binary expansion of m. Thus, we reach the fact that the partition C, D equals the partition X, Y, proving the uniqueness of such a partition.

5 Constructing a Square-free Sequence Over Three Symbols

Definition 6. A square free sequence is a sequence that contains no pattern of any length twice in a row. If X is a word of sequence values of length greater than or equal to 1 (we say $|X| \geq 1$), then a square-free sequence never contains that pattern XX, the concatenation of X with itself. It is clear that if we consider sequences of infinite length, then no sequence over an alphabet of size 1 can be square free, as the sequence must repeat the symbol in its alphabet twice in a row. It is almost as clear that one cannot construct a square-free sequence over an alphabet of size 2.

Theorem 4. There exists no infinite sequence of alphabet size 2 that is square free.

Proof. Without loss of generality, let the alphabet $\sum$ be defined as $\{0, 1\}$ so that $|\sum| = 2$. Assume for now that we start with a 0. In order to avoid an immediate square, we must then choose 1. We can't choose a 1 again so we choose 0. But now, we are stuck – if we choose a 0, the sequence is 0100, which contains a square in the last two digits. If we choose a 1, we have 0101, which is a square where the pattern X is 01. So, we cannot start with a 0. However, if we start with 1, we simply run into the same problem (the exact same argument suffices, with the values of the sequence toggled).

This development begs the question – how large must the alphabet of a sequence be in order to have infinite square free words? It turns out that the alphabet need only have one more symbol. Using the Thue-Morse sequence, we will be able to construct such a sequence with three symbols in its alphabet and that is infinite and square-free. First, we must prove a property of the TM sequence.

5.1 Overlap-free Property

Definition 7. An infinite sequence (or word) is overlap-free if there consists of no pattern of the form $aXaXa$, where a is a single term in the sequence (letter in the word), and X is a pattern of any number of letters (possibly 0).

Theorem 5. The Thue-Morse infinite word is overlap-free.

Proof. The following proof of Allouche and Shallit [1] contains 5 cases, but they are all short and simple. First, note that if the Thue-Morse sequence contains an overlap, then the infinite TM word can be represented as $UaXaXaV$, where U and X are patterns of length 0 or more, and V is a pattern of infinite length. Also, define the numbers k and m, where $k = |U|$ (the length of U) and $m = |aX|$ (the length of aX). The overlap that we assume exists means that $T_{k+j} = T_{k+j+m}$, for $0 \leq j \leq m$. Note that m must be at least one, even allowing for the case when $|X| = 0$, as $|a|$ always equals 1. Also note that no matter how complicated an infinite word's overlap is, there must be a smallest m value that yields an overlap. We will take our m value to be the smallest possible m, and find a contradiction. Then there is no smallest m value such that an overlap occurs and therefore no m value, no matter how large, such that an overlap occurs. This will prove the non-existence of overlaps.

Case 1: m and k are even

Let $m = 2m^*$ and $k = 2k^*$. Recall that we know from the overlap definition that $T_{k+j} = T_{k+j+m}$ for $0 \leq j \leq m$. Then, it is clear that $T_{k+2j^*} = T_{k+2j^*+m}$ for $0 \leq j^* \leq \frac{m}{2}$, where $j = 2j^*$ and when j^* is an integer. Since k is even, we also have $T_{2k^*+2j^*} = T_{2k^*+2j^*+2m^*}$ for $0 \leq j^* \leq m^*$. Recall from definition 1 that $T_n = T_{2n}$. Therefore, we can factor out the two and remove it, obtaining $T_{k^*+j^*} = T_{k^*+j^*+m^*}$ for $0 \leq j^* \leq m^*$. This is the condition for another (smaller) overlap, which contradicts the assumption that m was the smallest possible value that allowed an overlap.

Case 2: m is even and k is odd

Using the previous notation, $k = 2k^* + 1$. Similarly to before, we have $T_{2k^*+2j^*+1} = T_{2k^*+2j^*+2m^*+1}$ for $0 \leq j^* \leq m^*$. By definition 1, $T_{2\alpha+1} = \overline{T_\alpha}$. Application of this yields $\overline{T_{k^*+j^*}} = \overline{T_{k^*+j^*+m^*}}$. So, we have reached the following contradiction: $T_{k^*+j^*} = T_{k^*+j^*+m^*}$ for $0 \leq j^* \leq m^*$

Now, before looking at the last three cases, we need to define a new sequence, named b_n, where $b_n = (T_n + T_{n-1}) \pmod 2$ for $n \geq 1$. Notice that $b_{4n+2} = (T_{4n+2} + T_{4n+1}) \pmod 2$. Since the binary expansions of $4n+1$ and $4n+2$ have the same number of ones, $T_{4n+2} = T_{4n+1}$. Thus, b_{4n+2} equals either $(0+0)$ mod 2 or $(1+1) \pmod 2$, either of which are equal to 0. We will call this property 1. Reasoning similarly, we have $b_{2n+1} = 1$, as $T_{2n} = \overline{T_{2n+1}}$ (this means that b_{2n+1} equals $(0+1)$ (mod 2) or $(1+0) \pmod 2$, which both equal 1). This is property 2.

Case 3: m is odd and not less than 5

Because we know that $T_{k+j} = T_{k+j+m}$ for $0 \leq j \leq m$, we can use the definition of b_n to reach $b_{k+j} = b_{k+j+m}$ for $1 \leq j \leq m$. Since m is at least 5, we can choose a value for j such that the residue of $(k+j) \pmod 4$ is 2. (Since j ranges through at least 4 numbers, it has all possible residues, mod 4. Therefore, $k+j$ has all possible residues, mod 4, and a value of j exists such that the residue is 2). By property 1, $b_{k+j} = 0$ for this j value. Notice that $k+j+m$ is odd, simply because m is odd and $k+j$ is even, as their sum is congruent to 2 (mod 4). Then, by property 2, $b_{k+j+m} = 1$. However, we had earlier that $b_{k+j} = b_{k+j+m}$. So, we have arrived at a contradiction.

Case 4: m is odd and equals 3

If m is equal to three, then we can say that $b_{k+j} = b_{k+j+m}$ for $1 \leq j \leq 3$, for the same reasons as before. Because j will hold all but one possible residue (mod 4), we can be certain that $k+j$ will have a residue of either 2 or 3 (or possibly both), mod 4. If $k+j$ has a residue of 2 for some j, then the reasoning of the above case yields a contradiction. If the residue is three (mod 4), then the residue is also 1 (mod 2). Then, by property 2, $b_{k+j} = 1$. However, $k+j$ is odd as it is congruent to 3 (mod 4). Then, $k+j+3$ is even, and by property 1, $b_{k+j+3} = 0$. This is a contradiction as we knew earlier that $b_{k+j} = b_{k+j+m}$ for $1 \leq j \leq 3$.

Case 5: m **is odd and equals 1**

If an overlap occurs in this case, then $T_n = T_{n+1} = T_{n+2}$ for some n. However, if n is even and equals $2n^*$, then we have $T_{2n^*} = T_{2n^*+1}$ from the first equality, which is a contradiction. If n is odd and equals $2n^*+1$, then the second equality gives us the same contradiction: $T_{2n^*} = T_{2n^*+1}$.

This completes the proof that the Thue-Morse sequence is overlap-free. This yields an immediate consequence.

Corollary 1. The Thue-Morse sequence is cube-free and non-periodic.

Proof. If the TM sequence contained three identical terms in a row, or if the TM sequence was eventually periodic, then the TM sequence would have an overlap.

Now that the rather lengthy proof is behind us, we can quickly construct a square-free sequence over an alphabet of size three.

5.2 The Square-free Sequence

Definition 8. Define the sequence c_n, such that c_n is the number of 1's between the nth and $(n+1)$st zeroes. Note that this sequence starts with $n = 1$, as there is only a first zero, not a zeroth zero. The sequence c_n starts like: $2, 1, 0, 2, 0, 1, ...$ We know that this sequence contains only three symbols as there can be no number less than zero (obvious) or greater than three (as this would require the TM sequence to have an overlap). Now, to show that c_n is square-free, assume there is a square where the pattern P is repeated twice in a row. Denote the pattern of P to be $p_1 p_2 ... p_n$. Then, we know that at some point in the TM sequence, we have:

$$01^{p_1}01^{p_2}0...01^{p_n}01^{p_1}01^{p_2}0...01^{p_n}0$$

(The exponentiation of a term in the sequence simply means to concatenate it that number of times.) The only thing left to do to reach a contradiction is to note that this would constitute an overlap in the TM sequence, where in the above $UaXaXaV$ notation, U is everything before this subsequence (U could be of zero length), a is 0, X is $1^{p_1}01^{p_2}0...01^{p_n}$, and V is the infinite string after the subsequence. So, we have found a sequence of three symbols that is square-free, and interestingly enough, it is extremely closely related to the TM sequence.

6 A Final Application

Consider the following sequence of rational numbers, first considered in 1978 [10,12]:

$$\frac{1}{2}, \frac{\frac{1}{2}}{\frac{3}{4}}, \frac{\frac{\frac{1}{2}}{\frac{3}{4}}}{\frac{\frac{5}{6}}{\frac{7}{8}}} = \frac{\left(\frac{\left(\frac{1}{2}\right)}{\left(\frac{3}{4}\right)}\right)}{\left(\frac{\left(\frac{5}{6}\right)}{\left(\frac{7}{8}\right)}\right)}, ...$$

We will find what this infinite sequence converges to using the Thue-Morse sequence (Allouche and Shallit [1]). We can consider the given long division strings as being a single fraction with a single numerator and denominator consisting of different values. Let's call this simplified fraction the SSF, which stands for standard simplified fraction. For example, SSF_k, which equals the SSF of the k^{th} term equals $\frac{1*4}{2*3}$ for $k = 1$. Note that any SSF contains only one division sign. The first challenge is to find which numbers appear in the numerator and denominator of SSF_k. However, we can see that 1 always appears in the numerator and 2 always appears in the denominator. Then, the 3 and the 4 are in the opposite positions as the 1 and 2, due to the division sign. Similarly, the next group of numbers are in the opposite positions as the original four numbers. This process continues, and it is easy to see that definition 3 describes this pattern perfectly well. One could also consider a sequence of two symbols, $+$ and $-$. If we are looking at the k^{th} term of the original sequence of rational numbers, then there are 2^{k+1} numbers in the fraction, half of which are in the numerator (or denominator) of the SSF. If the sequence over the terms $+$ and $-$ assigns a $+$ to the integers that will end up being multiplied in the numerator of the SSF for a given term k and assigns a $-$ to integers that end up in the denominator, then it is clear that the sequence of $+$'s and $-$'s for the numbers 1 through 2^k is the opposite of the sequence values for the numbers $2^k + 1$ through 2^{k+1}, due to the flipping that occurs by the largest division sign. Note that, just like the Thue Morse sequence, this property (which is equivalent to definition 3) holds at all levels – not only along the single division sign of the SSF, but along every division sign in the term before simplification to a SSF. (Every division sign seperates 2^j integers for some j.) Therefore, the TM sequence can be used to determine which integers are in the numerator and denominator of the SSF; the k^{th} term of the above sequence of rational numbers is simply the following:

$$\prod_{n=0}^{2^{k+1}} (n+1)^{(-1)^{T_n}}.$$

One can also consider quotients of two numbers (instead of single numbers) to be the smallest units of the sequence terms. (Now, the $+$ and $-$ symbols would be assigned to rational numbers consisting of a single integer in the numerator and denominator, and would denote whether the fraction was to appear inverted or not in the SSF.) By making this change and by taking the limit, one can also see that the limit of the sequence of rational numbers is:

$$\prod_{n=0}^{\infty} \left(\frac{2n+1}{2n+2}\right)^{(-1)^{T_n}}.$$

We will now see that this infinite product converges to a non-zero finite number. First, take logs in order to convert into a sum.

$$log\left(\prod_{n=0}^{N} \left(\frac{2n+1}{2n+2}\right)^{(-1)^{T_n}}\right) = \sum_{n=0}^{N}(-1)^{T_n} log\left(\frac{2n+1}{2n+2}\right) = -\sum_{n=0}^{N}(-1)^{T_n}\left(-log\left(1-\frac{1}{2n+2}\right)\right)$$

Note that Dirichlet's test [5], tells us that the series converges, as the $-log\left(1-\frac{1}{2n+2}\right)$ term is decreasing to zero and the $(-1)^{T_n}$ term's partial sums are bounded in absolute value by 1 (as after any even number of terms, the same number of ones and zeroes have appeared in the TM sequence). So, we have the convergence of the sum. Say that it equals β. Then, our product equals e^{β}, which is finite and non-zero. Let $D = e^{\beta}$ be the quantity that we are considering. Then,

$$D = \prod_{n=0}^{\infty} \left(\frac{2n+1}{2n+2}\right)^{(-1)^{T_n}}.$$

Define

$$E = \prod_{n=1}^{\infty} \left(\frac{2n}{2n+1}\right)^{(-1)^{T_n}},$$

which converges due to the convergence of

$$-\sum_{n=1}^{\infty}(-1)^{T_n}\left(-log\left(1-\frac{1}{2n+1}\right)\right).$$

Then, we multiply to get:

$$DE = \frac{1}{2}\prod_{n=1}^{\infty}\left(\frac{n}{n+1}\right)^{(-1)^{T_n}}.$$

Seperation of the product into even and odd terms yields

$$DE = \frac{1}{2}\prod_{n=0}^{\infty}\left(\frac{2n+1}{2n+2}\right)^{(-1)^{T_{2n+1}}}\prod_{n=1}^{\infty}\left(\frac{2n}{2n+1}\right)^{(-1)^{T_n}}.$$

Then, since $T_n \neq T_{2n+1}$, we can see from this latest form that $DE = \frac{1}{2}D^{-1}E$, which implies that $D = \frac{\sqrt{2}}{2}$. So, we have reached the amusing result:

Theorem 6.

$$\frac{\frac{\frac{1}{2}}{\frac{3}{4}}}{\frac{\frac{5}{6}}{\frac{7}{\vdots}}} = \frac{\left(\frac{\left(\frac{1}{2}\right)}{\left(\frac{3}{4}\right)}\right)}{\frac{\left(\frac{\left(\frac{5}{6}\right)}{\left(\frac{7}{8}\right)}\right)}{\vdots}} = \prod_{n=0}^{\infty}(n+1)^{(-1)^{T_n}} = \prod_{n=0}^{\infty}\left(\frac{2n+1}{2n+2}\right)^{(-1)^{T_n}} = \frac{\sqrt{2}}{2}.$$

7 Generalizations

It should be mentioned that the Thue-Morse sequence can be generalized into other bases. The version of the sequence discussed was defined by the sum of the digits of the binary representation of an integer, mod 2. However, general TM sequence are defined as the sum of digits of the base k, mod m. Our special case where $k = m = 2$ was proven to be overlap-free, but this is not true for general TM sequences. In fact, Allouche and Shallit [2] proved that a Thue-Morse sequence is overlap-free if and only if $m \geq k$ (for $k \geq 2$ and $m \geq 1$). For further information on infinite products like the one discussed in the previous section, see [3,4,11]. Finally, the TM sequence can also be shown to be closely related to important features of the Mandelbrot set and to be tightly linked to the Koch curve [8].

8 Acknowledgements

The author is pleased to thank Professors Jeffrey Shallit (University of Waterloo, Ontario) and Ralph Greenberg (University of Washington, Seattle) for their prompt answers to various queries.

9 References

.[1] Allouche, J.-P. and Shallit, J. Automatic Sequences: Theory, Applications, Generalizations. Cambridge, England: Cambridge University Press, 2003.

.[2] J.-P. Allouche and J. Shallit, Sums of digits and overlap-free words, in preparation.

.[3] J.-P. Allouche and J. Shallit, Infinite Products Associated with Counting Blocks in Binary Strings, J. London Math. Soc. (1989) s2-39 (2): 193-204.

.[4] Allouche, J.-P. ”Series and Infinite Products Related to Binary Expansions of Integers.” 1992.

.[5] Folland, Gerald. Advanced Calculus. Upper Saddle River, NJ: Prentice-Hall, 2002. 303. Print.

.[6] Honsberger, Ross. Mathematical Diamonds. The Mathematical Association of America, 2003. 143-150. Print.

.[7] J. Lambek and L. Moser, On some two way classifications of integers, Canad. Math. Bull. 2 (1959), 85-89.

.[8] J. Ma, J. Holdener, When Thue-Morse meets Koch, Fractals 13 (2005) 191206.

.[9] Prouhet, E. "M*é*moir sur quelques relations entre les puissances des nombres." C. R. Adad. Sci. Paris S*é*r. 1 33, 225, 1851.

.[10] D. Robbins, Solution to Problem E 2692, Am. Math. Monthly 86 (1979), 394-5.

.[11] Shallit, J. O. "On Infinite Products Associated with Sums of Digits." J. Number Th. 21, 128-134, 1985.

.[12] D. R. Woods, Elementary Problem Proposal E 2692, Am. Math. Monthly 85 (1978), 48.