Exploring the Thue-Morse sequence

Adam Hutchings, Jake Roggenbuck

July 5, 2022

1 Introduction

The Thue-Morse sequence can be defined in many ways, but it is the sequence whose nth term is the number of ones in the binary representation of n. The sequence T_3 defined as the nth term being $T(3n)$ has some interesting properties that we will investigate.

2 No-2 Theorem

We define a *motif* in T_3 to be a sequence of the same digit not contained inside any larger motif. The authors observed that for the first $10^1 1$ terms of T_3, there are no motifs of length 2, which can be proven formally.

No-2 Theorem There are no motifs of length 2 in T_3.

Proof We begin by observing that a motif of zeros of length 2 is the sequence 00 surrounded by 1s, and a motif of ones of length 2 is the opposite. These translate to 1001 and 0110 in T_3, which must exist if the No-2 Theorem is false. Because we obtain the elements of T_3 from choosing every third element of the Thue-Morse sequence, the existence of 1001 or 0110 in T_3 implies the existence of $1XX0XX0XX1$ or $0XX1XX1XX0$ in the Thue-Morse sequence where X is either digit.

Lemma 2.1 The sequences 000 and 111 do not exist in the Thue-Morse sequence.

Proof The non-existence of repetitions of three in the sequence is a well-known fact, but may be established as follows: jumping from $2n$ to $2n+1$ changes a 0 to a 1 and nothing else, which means that T_{2n} and T_{2n+1} are opposite. Therefore, the entire Thue-Morse sequence consists of repetitions of 01 and 10, which allows for no 000 or 111.

Lemma 2.1 allows for a computer program to comb through all ten-digit sequences of zeros and ones, with the constraints that they follow the pattern $0XX1XX1XX0$ and do not contain 000 or 111. This does not check any occurrences of 1001 in T_3, but this is not necessary.

Lemma 2.2 A sequence X exists in the Thue-Morse sequence *iff* the sequence X_i also exists, where X_i is X with 0 and 1 interchanged.

Proof Such a sequence X will occur entirely inside the first n digits of the sequence, where n is an arbitrarily large power of 2. The *next* n digits are the same as the first n with 0 and 1 interchanged, so the existence of X in the first block is equivalent to the existence of X_i in the second.

The only ten-digit sequences that satisfy these constraints are 0011001010, 0011001100, 0011011010,0101001010, 0101001100, 0101011010, 0101101010, and 0101101100.

Lemma 2.3 If n is some power of 2 and the sequence X is not more than n digits long, then if X does not occur in the first $8n$ digits of the Thue-Morse sequence, it never occurs.

Proof Call the first n digits of the Thue-Morse sequence A, and A_i (as defined above) B. Then by the same logic as the proof of Lemma 2.2, the first $8n$ digits are $ABBABAAB$. As the sequence n is no more than n digits long, any occurrence of it is either entirely within one n-long block or straddling two. Therefore, the only possible "environments" in which the pattern may occur are A, B, AA, AB, BA, or BB. All of these environments occur in $ABBABAAB$, so if the pattern never exists there, it never exists anywhere.

Now, Lemma 2.3 means we only need to check the first 128 digits (as the 10-digit sequences are less than 16 long, so checking 128 is sufficient.) Checking using a computer program rules these out, so the No-2 Theorem is proven.