Introduction

This Power Round develops the many and varied properties of the Thue-Morse sequence, an infinite sequence of 0s and 1s which starts $0, 1, 1, 0, 1, 0, 0, 1, \ldots$ and appears in a remarkable number of different contexts in recreational and research mathematics. We will see applications to geometry, probability, game theory, combinatorics, algebra, and fractals. Nevertheless, we won't even come close to exhausting the amusing and useful properties of this sequence, some of which require mathematics beyond our scope to discuss.

Remark: Regardless of which problem you decide to work on, it is recommended that you read Problem 1 first to become familiar with the definitions.

Remark 2: The following problems rely heavily on the technique of proof by induction. If you are not yet comfortable with induction, we have copies of an introduction available for you to consult—ask your proctor.

Defining the Thue-Morse sequence

The first sign that there's something special about the Thue-Morse sequence is that it's hard to make up your mind about how to define it, because there are numerous very different-looking definitions which all turn out to be equivalent. In this problem, we work through a few of these definitions and determine that each of them gives the same result. We refer to the nth term of the Thue-Morse sequence by t_n, starting with $t_0, t_1, t_2, \ldots$.

1. (a) **[3]** Our first definition is a simple recursive one. The zeroth term of the Thue-Morse sequence is $t_0 = 0$. For n a nonnegative integer, after the first 2^n terms of the Thue-Morse sequence (including the zeroth term) have been specified, construct the next 2^n terms by taking the first 2^n terms, replacing each 0 by a 1, and replacing each 1 by a 0 (simultaneously). (This is called "bitwise negation".) Therefore, we have $t_1 = 1$, and the next two terms are $t_2 = 1, t_3 = 0$. The zeroth through fifteenth terms (leaving out the commas, as we will often do for convenience) are 0110100110010110.

   Write down (no justification required) the 16th through 31st terms.

   (b) **[6]** Our second definition is direct. The Thue-Morse sequence is the sequence $\{t_n\}$ ($n = 0, 1, \ldots$) where t_n is 1 if the number of ones in the binary (base-2) expansion of n is odd and 0 if the number of ones in the binary expansion of n is even. For example, 5 is 101_2 in base 2, which has two ones, so $t_5 = 0$.

   Prove that this definition gives the same sequence as the one from part (a).

   (c) **[6]** Our third definition is recursive again, but uses a different recursion. The Thue-Morse sequence is the sequence $\{t_n\}$ satisfying $t_0 = 0$, $t_{2n} = t_n$, and $t_{2n+1} = 1 - t_n$.

   Prove that this definition is equivalent to either of the first two definitions.

   (d) **[6]** Our fourth definition is by a certain algorithm (known as a Lindenmeyer system). We start with the single digit 0 (call this stage zero). At each stage, we take the digits we already have, replace each 0 by a 01, and replace each 1 by a 10 (simultaneously). So stage one is 01, stage two is 0110, and so on. The Thue-Morse sequence is the sequence $\{t_n\}$ whose first 2^n terms are the digits from stage n.

   Prove that this definition is equivalent to any of the first three definitions. (Note that as stated, it is not clear that this definition is even coherent, since it redefines each term over and over again. Your job is to show that it nevertheless uniquely defines each term as the corresponding term of the Thue-Morse sequence as given by parts (a)-(c).)

Solution to Problem 1:

(a) 1001011001101001.

(b) Let $\{u_n\}$ be the sequence given by this definition and $\{t_n\}$ be the sequence from part (a). We proceed by induction. For the base case, we see that $u_0 = 0 = t_0$. Assume that $u_i = t_i$ for i from 0 to $2^n - 1$; then we claim that also $u_i = t_i$ for i from 2^n to $2^{n+1} - 1$. This is because if $0 \le i \le 2^n - 1$, then i has at most n digits in base 2, so the binary expansion of $i + 2^n$ is the same as the binary expansion of i except with an extra 1 and maybe some 0s attached to the left from the 2^n. Therefore the parity of the number of 1s in $i + 2^n$ base 2 is always different from the parity of the number of 1s in i base 2, so that $u_{i+2^n} = 1 - u_i$. But also $t_{i+2^n} = 1 - t_i$ by definition. By the inductive hypothesis, we have $u_{i+2^n} = 1 - u_i = 1 - t_i = t_{i+2^n}$. This completes the induction.

(c) Let $\{u_n\}$ be the sequence given by this definition and $\{t_n\}$ be the sequence from parts (a)-(b). We proceed by induction. For the base case, we see that $u_0 = 0 = t_0$. Assume that $u_i = t_i$ for i from 0 to $2^n - 1$; then we claim that also $u_i = t_i$ for i from 2^n to $2^{n+1} - 1$. This is because if $0 \le j \le 2^n - 1$, then $u_{2j} = u_j$, $u_{2j+1} = 1 - u_j$. But we also have $t_{2j} = t_j$ because the binary expansion of $2j$ is the same as the binary expansion of j except with a 0 attached to the right, so they have the same number of 1s, and $t_{2j+1} = 1 - t_j$ because the binary expansion of $2j + 1$ is the same as the binary expansion of j except with a 1 attached to the right, so the numbers of 1s in the two expansions always have different parities. By the inductive hypothesis, $t_{2j} = t_j = u_j = u_{2j}$, $t_{2j+1} = 1 - t_j = 1 - u_j = u_{2j+1}$. This completes the induction.

(d) We prove by induction on n that if we run $t_0 \cdots t_{2^n-1}$ through one round of this algorithm (which we will call F), the result is $t_0 \cdots t_{2^{n+1}-1}$. For the base case, we see that $F(0) = 01$. Assume that $F(t_0 \cdots t_{2^{n-1}-1}) = t_0 \cdots t_{2^n-1}$. Let G be the bitwise negation function. Note that $F(G(0)) = F(1) = 10 = G(01) = G(F(0))$ and similarly $F(G(1)) = G(F(1))$, so in general $F(G(x)) = G(F(x))$ for a string x of 0s and 1s. Since $G(t_0 \cdots t_{2^{n-1}-1}) = t_{2^{n-1}} \cdots t_{2^n-1}$ and $G(t_0 \cdots t_{2^n-1}) = t_{2^n} \cdots t_{2^{n+1}-1}$, we conclude that

$$\begin{aligned} F(t_{2^{n-1}} \cdots t_{2^n-1}) &= F(G(t_0 \cdots t_{2^{n-1}-1})) \\ &= G(F(t_0 \cdots t_{2^{n-1}-1})) \\ &= G(t_0 \cdots t_{2^n-1}) \\ &= t_{2^n} \cdots t_{2^{n+1}-1} \end{aligned}$$

by the inductive hypothesis. This completes the induction.

Now we derive a few simple properties of the Thue-Morse sequence, just to play with it some more.

2. (a) **[5]** Prove that the string $t_0 t_1 \cdots t_{2^{2n}-1}$ is a palindrome for all $n \ge 0$. (Recall that a palindrome is a string of digits which reads the same forward and backward.)

(b) Let A be the set of all nonnegative integers n such that $t_n = 0$. Let $n \oplus m$ denote the *binary xor* of n and m. (To compute the binary xor of n and m, we write both n and m in binary, then add them without carrying. For example, if $n = 5$ and $m = 13$, then $n = 101_2$ and $m = 1101_2$, so $n \oplus m = 1000_2 = 8$.)

(i.) **[1]** Compute $14 \oplus 23$.

(ii.) **[5]** Prove that if n and m are both in A, then $n \oplus m$ is also in A.

(c) **[6]** Prove that given any finite string $X = t_a t_{a+1} \cdots t_b$ of consecutive terms from the Thue-Morse sequence, there exists a number n_X such that *every* string of n_X consecutive terms $t_{k+1}t_{k+2}\cdots t_{k+n_X}$ from the sequence must contain X.

(d) **[6]** Given a finite or infinite string T of 0s and 1s, let $f(T)$ be the string created by simultaneously replacing each 0 by a 01 and each 1 by a 10. For example, if $T = 001$, then $f(T) = 010110$. Note that we previously saw this procedure in problem 1, part d. A *fixed point* of f is an infinite string T such that $f(T) = T$. Prove that f has exactly two fixed points: the Thue-Morse sequence $\{t_n\}$, and its bitwise negation (meaning the sequence constructed from $\{t_n\}$ by replacing each 0 with a 1 and each 1 with a 0).

Solution to Problem 2:

(a) We proceed by induction. For the base case, we see that this is true for $n = 0$, when the string is the single character t_0. Assume that $t_0 \cdots t_{2^{2n-2}-1}$ is a palindrome. Let G be the bitwise negation function. Then we have

$$\begin{aligned} t_{2^{2n-2}} \cdots t_{2^{2n-1}-1} &= G(t_0 \cdots t_{2^{2n-2}-1}) \\ &= t_{2^{2n-1}} \cdots t_{2^{2n-1}+2^{2n-2}-1} \end{aligned}$$

and

$$\begin{aligned} t_0 \cdots t_{2^{2n-2}-1} &= G(t_{2^{2n-2}} \cdots t_{2^{2n-1}-1}) \\ &= t_{2^{2n-1}+2^{2n-2}} \cdots t_{2^{2n}-1}. \end{aligned}$$

By the inductive hypothesis, each of these strings is a palindrome (since bitwise negation takes palindromes to palindromes). But it is clear that a sequence of four palindromes, in which the inner two are the same and the outer two are the same, is also a palindrome. This completes the induction.

Alternate solution: we can prove this directly from 1(b). Checking that $t_0 \cdots t_{2^{2n}-1}$ is a palindrome means checking that $t_i = t_{2^{2n}-1-i}$. But $2^{2n} - 1$ in binary is a string of $2n$ ones, so when we subtract the binary representation of i, we find that if i has k ones in its binary representation, then $2^{2n} - 1 - i$ has $2n - k$ ones. But k and $2n - k$ are of the same parity, so we are done.

(b) (i) $14 = 1110_2$ and $23 = 10111_2$, so $14 \oplus 23 = 11001_2 = 25$.

(ii) We showed in Problem 1, part b that $t_n = 0$ if and only if the number of ones in the binary expansion of n is even. Suppose n has $a(n)$ ones in its binary expansion and m has $a(m)$ ones, where $a(m), a(n)$ are both even since $t_n = t_m = 0$. If k of these ones are in the same digit place in both binary expansions, then $n \oplus m$ has $a(n) + a(m) - 2k$ ones in its binary expansion, since we "lose" 2 ones every time we add $1 + 1$ in one of the places in the expansion where the result is 0. But $a(n) + a(m) - 2k$ is a sum of three even numbers, is therefore even, and so $t_{n \oplus m} = 0$.

(c) Let $Y = t_0 \cdots t_{2^n-1}$ be the shortest string of consecutive terms from the Thue-Morse sequence which starts with t_0, has length a power of 2, and contains X. As in part a, we see that also $Y = t_{2^{n+1}+2^n} \cdots t_{2^{n+2}-1}$, and by the same reasoning we conclude that furthermore $Y = t_{2^{n+2}+2^n} \cdots t_{2^{n+2}+2^{n+1}-1} = t_{2^{n+2}+2^{n+1}} \cdots t_{2^{n+2}+2^{n+1}+2^n-1}$. In fact, if we split the Thue-Morse sequence into blocks of size 2^{n+2}, the string Y appears twice in every block, once in the first half and once in the second half (this is a quick proof by induction as usual). By the Pigeonhole Principle, every sequence of 2^{n+2} consecutive terms overlaps at least one of these blocks in at least half of its terms, and therefore contains at least one copy of Y. So we are done.

(Note: another way to think about this is that the Thue-Morse sequence consists of copies of Y and its bitwise negation $G(Y)$ arranged in a bigger copy of the Thue-Morse sequence, which is cube-free; hence every sequence of three such copies contains a copy of Y.

More concretely, Y is a block of size 2^n. Note that if we divide the Thue-Morse sequence into blocks of size $2 \cdot 2^n$, by Pigeonhole, a consecutive sequence of $4 \cdot 2^n$ terms must contain at least 1 of these blocks, running from $k \cdot 2^{n+1}$ to $(k+1)2^{n+1} - 1$ for some nonnegative integer k. Recalling the construction in 1(d) using the Lindenmeyer function F, we note that the subsequence $t_{k2^{n+1}} \dots t_{(k+1)2^{n+1}-1}$ comes from $F^{n+1}(t_k)$. But this is $F^n(F(t_k)) = F^n('01' \text{ or } '10')$, and $F^n(0) = Y$.)

Alternate solution: We use the binary representation in 1(b). Let n denote the number of binary digits in b. We claim that any consecutive sequence of $8 \cdot 2^n$ terms $t_c \dots t_{c+2^{n+3}}$ contains a copy of $t_a \dots t_b$. Indeed, let $c' := c$ rounded up to the nearest multiple of $2 \cdot 2^n$; then $t_{c'+a} \dots t_{c'+b}$ is a copy of the original sequence is c' has an even number of 1's in the binary expansion. If not, then $c' + 2^n$ has an even number of 1's, and hence $t_{c'+2^n+a} \dots t_{c'+2^n+b}$ works.

(d) We already essentially argued that $\{t_n\}$ is indeed a fixed point of this operation. To be precise, the claim that T is a fixed point means that if $f(T) = s_0 s_1 \cdots$, then $t_i = s_i$ for all i. But we proved in problem 1 part d that $f(t_0 \cdots t_{2^n-1}) = t_0 \cdots t_{2^{n+1}-1}$, which shows that $t_i = s_i$ for $0 \le i \le 2^{n-1}$. Since n is arbitrary, T is indeed a fixed point, and so is its bitwise negation by the same proof.

To prove that there are no others, it suffices to note that if S is a fixed point which begins with 0, then since $S = f(S)$, also $S = f^n(S)$ (where f^n means f composed with itself n times), the first 2^n terms of S must equal $f^n(0)$, which is just the first n terms of the Thue-Morse sequence. Since n is arbitrary, S must be exactly $\{t_n\}$. The same argument for the bitwise negation of $\{t_n\}$ holds if S starts with 1.

(It is also possible to use 1(d) for the second part as well.)

Greedy Galois Games

Time for some probability and game theory. Alice and Bob are in a duel where in each round (beginning with round 0), one duelist fires a shot at the other, hitting them with a success probability of p. The first person to fire a successful shot wins. They want to choose the shooter each round in a way that's fair—just switching back and forth after every shot wouldn't be fair, since we can see intuitively that whoever goes first is more likely to win. Also, they're both terrible at aiming, so p is very low, though positive. What do they do?

They come up with the following idea: Alice shoots first. Then, Bob shoots as many times as is necessary for his win probability to meet or exceed that of Alice's win probability so far. Then, Alice starts shooting again, again taking as many turns as is necessary for her win probability to meet or exceed that of Bob's win probability. And so on (if at any point, they have the same probability of winning, we let the person who was not shooting in the previous round shoot in the next round).

For example, suppose $p = 1/3$. Alice shoots during round 0, after which her win probability is 1/3 and Bob's win probability is 0. Bob shoots during round 1. For Bob to win during round 1, Alice has to miss in round 0, which happens with probability 2/3, and Bob has to hit in round 1, which happens with probability 1/3. So after round 1, Bob's win probability is $(2/3)(1/3) = 2/9$, which is still less than Alice's win probability of 1/3. Therefore, Bob shoots again in round 2. By the same logic, his overall win probability after round 2 is $(2/3)(1/3) + (2/3)(2/3)(1/3) = 10/27$, which is now higher than 1/3. So Alice gets to shoot in round 3. And so on.

Let $P(A)$ be Alice's overall win probability after a given round, and $P(B)$ be Bob's win probability. We summarize the above information in the following table:

Round #	Shooter	$P(A)$	$P(B)$
0	Alice	1/3	0
1	Bob	1/3	2/9
2	Bob	1/3	10/27
3	Alice	?	10/27
4	?	?	?

3. (a) (i.) **[2]** Fill in the question marks in the above table (no justification required).

   (ii.) **[3]** Fill in the same table for $p = 1/4$ instead of 1/3 (no justification required).

   (b) **[6]** Let $q = 1 - p$. Let $\{a_n\}$ be the sequence such that $a_n = -1$ if Alice shoots in round n and $a_n = 1$ if Bob shoots in round n. Let $P(A_n)$ be Alice's overall win probability after round n, and $P(B_n)$ Bob's overall win probability after round n. Finally, let

$$f_n(x) = a_n \left(\sum_{j=0}^{n} a_j x^j \right).$$

   Prove that

$$a_{n+1} = \begin{cases} -a_n & \text{if } f_n(q) \geq 0, \\ a_n & \text{otherwise.} \end{cases}$$

   (c) **[3]** Prove that regardless of the value of p, we always have $a_0 = -1, a_1 = 1, a_2 = 1$.

   (d) **[3]** Determine, with proof, all values of p such that $a_3 = -1$.

Solution to Problem 3:

(a) (i.)

Round #	Shooter	$P(A)$	$P(B)$
0	Alice	1/3	0
1	Bob	1/3	2/9
2	Bob	1/3	10/27
3	Alice	35/81	10/27
4	Bob	35/81	106/243

(ii.)

Round #	Shooter	$P(A)$	$P(B)$
0	Alice	1/4	0
1	Bob	1/4	3/16
2	Bob	1/4	21/64
3	Alice	91/256	21/64
4	Bob	91/256	417/1024

(b) Note that

$$P(A_n) = \sum_{j:a_j=-1} pq^j,$$

since for each given round that Alice participates in, there is a probability q^j of reaching round j, and probability p that Alice wins in that round. Similarly,

$$P(B_n) = \sum_{j:a_j=1} pq^j.$$

Hence, we can write

$$f_n(q) = \frac{a_n}{p}\left(P(B_n) - P(A_n)\right).$$

If $f_n(q) < 0$, then either $a_n < 0$ and $P(B_n) > P(A_n)$ or $a_n > 0$ and $P(B_n) < P(A_n)$. In either case, the person who just shot has a lower probability of winning, so they should continue shooting, so $a_{n+1} = a_n$. Otherwise, the person who just shot has a higher (or equal) probability of winning, so the other person should start shooting, so $a_{n+1} = -a_n$.

(c) Alice always shoots first, so $a_0 = -1$ by default. $P(A_0) = p > P(B_0) = 0$, so Bob must shoot in round 1, so $a_1 = 1$. $P(A_1) = p > P(B_1) = pq$ since $q < 1$, so Bob must shoot again in round 2, so $a_2 = 1$.

(d) Note $P(A_2) = p$ and $P(B_2) = p(q+q^2)$. We have $a_3 = -1$ if and only if $P(A_2) \le P(B_2)$, so we need $q + q^2 \ge 1$, which is true if and only if $q \ge \frac{-1+\sqrt{5}}{2}$ since $q > 0$. Hence, we require $p \le 1 - \frac{-1+\sqrt{5}}{2} = \frac{3-\sqrt{5}}{2}$.

Our goal is now to prove that as p gets close to 0, or equivalently as q gets close to 1, the pattern of who shoots who becomes more and more like the Thue-Morse sequence, in the following sense. Recall that we define a_n to be -1 if Alice shoots in round n and 1 if Bob shoots in round n, and that $\{t_n\}$ is the Thue-Morse sequence. Let $\{t'_n\}$ be the sequence such that $t'_n = -1$ if $t_n = 0$ and $t'_n = 1$ if $t_n = 1$. That is, $\{t'_n\}$ is basically also the Thue-Morse sequence, just using -1 and 1 instead of 0 and 1, since that's more convenient for our current application. We're going to show that as p gets close to 0, more and more of the first few terms of $\{a_n\}$ equal the first few terms of $\{t'_n\}$.

4. (a) [**8**] Prove that for each $n \in \mathbb{N}$, there is an $\epsilon > 0$ such that the sequence $a_0, a_1, \ldots, a_n$ is the same for all $q \in (1-\epsilon, 1)$. Intuitively, this shows that as the success probability p nears zero, more and more of the first few terms of a_n stabilize and become fixed. (Hint: start with your solution to Problem 3).

(b) (i) [**3**] Prove that for any m, we have $\sum_{i=0}^{2m+1} t'_i = 0$.

(ii) [**7**] Suppose that there exists $\epsilon > 0$ such that for all $q \in (1-\epsilon, 1)$, $a_i = t'_i$ for $0 \le i \le 2m$. Prove that then there is an $\epsilon' > 0$ such that $a_{2m+1} = -a_{2m}$ for all $q \in (1-\epsilon', 1)$.

(c) [**6**] Suppose that there exists $\epsilon > 0$ such that for all $q \in (1-\epsilon, 1)$, $a_i = t'_i$ for $0 \le i \le 2m+1$. Prove that when $q \in (1-\epsilon, 1)$, $f_{2m+1}(q) = (q-1)f_m(q^2)$.

(d) [**6**] Prove that for each $n \in \mathbb{N}$, there is an $\epsilon > 0$ such that the sequence $a_0, a_1, \ldots, a_n$ is the same as the sequence $t'_0, t'_1, \ldots, t'_n$ for all $q \in (1-\epsilon, 1)$. (This demonstrates the claim we made in the paragraph before this problem.)

Solution to Problem 4:

(a) We induct on n. The base cases $n = 0, 1, 2, 3$ were shown in Problem 3. Now assume for some $n \in \mathbb{N}$ and $\epsilon > 0$, we have that $a_0, \ldots, a_n$ is the same sequence for all $q \in (1-\epsilon, 1)$. We want to find $\epsilon' > 0$ such that $a_0, \ldots, a_{n+1}$ is the same sequence for all $q \in (1-\epsilon', 1)$. Recall from problem 4 that

$$a_{n+1} = \begin{cases} -a_n & \text{if } f_n(q) \geq 0, \\ a_n & \text{otherwise.} \end{cases}$$

If the polynomial $f_n(q)$ does not have a zero in the interval $(1-\epsilon, 1)$, then a_{n+1} is the same for all $q \in (1-\epsilon, 1)$, and so we simply can set $\epsilon' = \epsilon$. Otherwise, let r denote the largest number in the interval $(1-\epsilon, 1)$ such that $f_n(r) = 0$. Then, a_{n+1} is the same for all $q \in (r, 1)$, so we set $\epsilon' = 1 - r > 0$.

(b) (i) We show that for any m, $t'_{2m} = -t'_{2m+1}$. The claimed identity follows trivially. From the third definition of the Thue-Morse sequence, we have that $t_{2m} = t_m$ and $t_{2m+1} = 1 - t_m$. Hence, $t'_m = 1 \implies t'_{2m} = 1, t'_{2m+1} = -1$ and $t'_m = -1 \implies t'_{2m} = -1, t'_{2m+1} = 1$. Either way, $t'_{2m} = -t'_{2m+1}$.

(ii) Note that $f_{2m}(1) = a_{2m} \sum_{j=0}^{2m} a_j = a_{2m} \cdot a_{2m} = 1$, since $\sum_{j=0}^{2m-1} a_j = \sum_{j=0}^{2m-1} t'_j = 0$. Let r be the largest number less than 1 such that $f_{2m}(r) = 0$ (or $-\infty$ if $f_{2m}(x)$ is always positive for $x < 1$), and let $\epsilon' = \min(\epsilon, 1-r)$. Clearly $\epsilon' > 0$. Since $1 - \epsilon' \geq 1 - \epsilon$, we have that for all $q \in (1-\epsilon', 1)$, $a_i = t'_i$ for $0 \leq i \leq 2m$. Moreover, for q in that interval, $f_{2m}(q) \geq 0$, so $a_{2m+1} = -a_{2m}$.

(c) We can rewrite

$$f_{2m+1}(q) = a_{2m+1} \sum_{i=0}^{m} (a_{2i} q^{2i} + a_{2i+1} q^{2i+1}) = a_{2m+1} \sum_{i=0}^{m} (a_{2i} + q a_{2i+1}) q^{2i},$$

by grouping the terms of $f_{2m+1}(x)$ into pairs. For $q \in (1-\epsilon, 1)$, this becomes

$$a_{2m+1} \sum_{i=0}^{m} (t'_{2i} + q t'_{2i+1}) q^{2i} = a_{2m+1} \sum_{i=0}^{m} (t'_{2i} - q t'_{2i}) q^{2i} = a_{2m+1} \sum_{i=0}^{m} t'_{2i} (1-q) q^{2i}.$$

Finally, definition 3 of the Thue-Morse sequence tells us that $t'_{2i} = t'_i$ for all i, so this equals

$$a_{2m+1}(1-q) \sum_{i=0}^{m} t'_i q^{2i} = (a_{2m+1}/a_m)(1-q) a_m \sum_{i=0}^{m} t'_i q^{2i} = (q-1) f_m(q^2),$$

as desired.

(d) We induct on n. The base cases $n = 0, 1, 2, 3$ were shown in Problem 3. Now assume that there exists some $\epsilon > 0$ such that for all $q \in (1-\epsilon, 1)$, $a_i = t'_i$ for all $0 \leq i \leq n$. There are two cases. First, assume n is even, i.e. $n = 2m$ for some integer m. By part (b), we know that there exists ϵ' such that $a_{n+1} = a_{2m+1} = -a_{2m}$ for all $q \in (1-\epsilon', 1)$. Since $t'_{2m+1} = -t_{2m}$, this exactly tells us that for all $q \in (1-\epsilon', 1)$, $a_i = t'_i$ for all $0 \leq i \leq 2n+1$, as desired.

Now consider the case when n is odd, i.e. $n = 2m+1$ for some integer m. Let $\epsilon' = 1 - \sqrt{1-\epsilon}$. Note that $0 < \epsilon' < \epsilon$, since $1 - \epsilon' = \sqrt{1-\epsilon} > 1 - \epsilon$, as squaring a positive number less than one makes it smaller. For all $q \in (1-\epsilon', 1)$, we have that $f_m(q)$ and $f_m(q^2)$

have the same sign, since $q \in (1-\epsilon', 1) = (\sqrt{1-\epsilon}, 1) \implies q^2 \in (1-\epsilon, 1)$, and so the sequences $\{a_i\}$ for q and q^2 are identical. Part (c) tells us that $f_{2m+1}(q)$ has the opposite sign as $f_m(q^2)$ and $f_m(q)$, since $q-1 < 0$. So, $a_{2m+2} = a_{2m+1}$ if and only if $a_{m+1} = -a_m$. Definition 3 of the Thue-Morse sequence tells us that $t'_{2m+1} = -t'_m$, and $t'_{2m+2} = t'_{m+1}$. Hence, if $a_{m+1} = -a_m$, then $a_{2m+2} = a_{2m+1} = t'_{2m+1} = -t'_m = t'_{m+1} = t'_{2m+2}$; and if $a_{m+1} = a_m$, then $a_{2m+2} = -a_{2m+1} = -t'_{2m+1} = t'_m = t'_{m+1} = t'_{2m+2}$. Either way, $a_{n+1} = a_{2m+2} = t'_{n+1} = t'_{2m+2}$, as desired.

Pattern avoidance

Now we develop and prove some more complicated but really cool properties of the Thue-Morse sequence. The goal of the next problem is to prove that no string of consecutive terms in the Thue-Morse sequence repeats itself three times consecutively. That is, the Thue-Morse sequence contains no *cubes*, where a cube is a nonempty string of consecutive terms which looks like www, where w is any string of 0s and 1s (for example, 001001001 is a cube with $w = 001$). As in some previous problems, we will leave out the commas between terms for convenience.

5. (a) **[3]** Of course, the simplest cubes are 000 and 111. Prove directly that in the Thue-Morse sequence, there are never three consecutive 0s or three consecutive 1s. (You may leave this part blank and receive full credit for it, but *only* if you receive full credit on the entire rest of this problem.)

   (b) We define an *overlapping factor* to be a nonempty string x of consecutive terms which begins with a string w of length shorter than x, and ends with the same string w, such that the two occurrences of w overlap in at least one term. For example, $x = 11011011$ is an overlapping factor because it both begins and ends with $w = 11011$, and the two instances of 11011 overlap by two terms (the middle two 1s).

   (i.) **[3]** Prove that if a sequence contains a cube, then it also contains an overlapping factor.

   (ii.) **[8]** Prove that if a sequence contains an overlapping factor, then it also contains an overlapping factor of the form $avava$, where a is a single term and v is a (possibly empty) string of terms.

   (c) **[5]** Suppose that $x = a_0 a_1 \cdots a_{2n-1}$ where each a_i is either 0 or 1 and each string $a_{2i}a_{2i+1}$ is either 01 or 10. Prove that it is not possible to write $0x0$ or $1x1$ in the form $b_0 b_1 \cdots b_{2n+1}$ where each b_j is either 0 or 1 and each string $b_{2i}b_{2i+1}$ is either 01 or 10.

   (d) Given a string T of 0s and 1s, let $f(T)$ be the function from problem 2, part d—that is, the string created by simultaneously replacing each 0 by a 01 and each 1 by a 10.

   (i.) **[6]** Suppose $f(T) = xavavay$ where a is a single term (0 or 1) and x, v, y are strings of 0s and 1s. Prove that v consists of an odd number of terms.

   (ii.) **[7]** Prove that if $f(T)$ contains an overlapping factor, then T also contains an overlapping factor.

   (iii.) **[3]** Prove that the Thue-Morse sequence contains no overlapping factors, and therefore no cubes.

Solution to Problem 5:

(a) We use the definition of the Thue-Morse sequence given in Problem 1, part d. We start with 0 and generate the sequence using the following replacement rule: $0 \to 01, 1 \to 10$. Notice that every 0 is adjacent to at least one 1, and every 1 is adjacent to at least one

0. This means that such simple cubes as 111 and 000 cannot appear in the sequence because they would require the middle 1 (or 0) to be without an adjacent 0 (or 1). We conclude that the Thue-Morse sequence does not contain three consecutive 0s or 1s.

(Note: there are many other ways to argue this. We can make the same conclusion from 1(b), or by induction from 1(a).)

(b) (i) If the sequence contains a cube, then there exists a string w such that www appears in the sequence. We can define the nonempty string x to be $x = ww$. We see that x must be an overlapping factor of the sequence, because the string www contains two instances of ww that overlap exactly in w. Thus, the presence of a cube implies the presence of an overlapping factor.

(ii) Let x be our overlapping factor containing two overlapping instances of w. Let n be the length of w, and let the overlap between w and itself start after the first k characters of x (i.e. x has length $n + k$). Finally, let x_i denote the i-th character of x.

Note that $x_i = x_{k+i}$ for all $i = 1, 2, \ldots, n$, since x contains w overlapping itself shifted by k characters. In particular, $x_1 = x_{k+1} = x_{2k+1}$, and the substring of $x_2x_3 \cdots x_k$ equals $x_{k+2}x_{k+3} \cdots x_{2k}$. So, we set $a = x_1$ and $v = x_2x_3 \cdots x_k$.

(c) Note that if a string can be written where each string $b_{2i}b_{2i+1}$ is either 01 or 10, then it contains equal numbers of 0's and 1's. x fits this criterion, and therefore contains exactly n 0's and n 1's. Therefore, $0x0$ and $1x1$ both have different numbers of 0's and 1's, so they cannot be written in the desired form.

(Note: there are other ways to argue this, e.g. by induction on the length of the sequence, or directly from writing $b_0 \cdots b_{2n+1} = 0a_0 \cdots a_{2n-1}0$ and concluding $b_0 = 0, b_1 = 1 = a_0, a_1 = 0 = b_2$, etc., up to $b_{2n+1} = 1$.)

(d) (i) First we note that $f(T)$ is even, since it has twice the length of T. Since $avava$ must have odd length, exactly one of x and y must have even length. We thus have two cases:

- Case 1: The length of x is odd and y is even, so $xa, vava, y$ consist of 10s and 01s.
- Case 2: The length of y is even and x is odd, so $x, avav, ay$ consist of 10s and 01s.

Now, assume that v has even length. Then, the length of ava must be even as well. In each of the cases above, v and ava can both be written as consecutive 10 and 01 terms. This stands in contradiction of the statement we proved in (c). Thus, we conclude that v must have odd length.

(Alternatively, $vava$ or $avav$ consisting of 10s and 01s implies that $vava$ or $avav$ has equal numbers of zeros and ones, hence that va or av has equal numbers of zeros and ones, hence that va or av is of even length, hence that v is of odd length.)

(ii) In part (b), we showed that if $f(T)$ has an overlapping factor, then it has an overlapping factor of the form $avava$, so that we can write $f(T) = xavavay$. Since v has odd length, we know that either va or av can be written as 10s and 01s. These correspond to the above two cases in part (i).

- Case 1: We can write $T = rsst$ such that $f(r) = xa, f(s) = va, f(t) = y$. Thus, since the strings r and s end with the same letter $\bar{a}$, we can write them as $r = r'\bar{a}, s = s'\bar{a}$. Thus, $T = r'\bar{a}s'\bar{a}s'\bar{a}t$ contains an overlapping factor of the form $w = \bar{a}s'\bar{a}$.
- Case 2: We can write $T = rsst$ such that $f(r) = x, f(s) = av, f(t) = ay$. Since the strings s and t start with the same letter a, we have $s = \bar{a}s', t = \bar{a}t'$. Thus, $T = r\bar{a}s'\bar{a}s'\bar{a}t'$ contains the overlapping factor $w = \bar{a}s'\bar{a}$.

We've shown that if $f(T)$ contains an overlapping factor, then T also contains an overlapping factor, completing our proof.

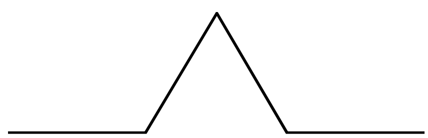
(iii) We proceed by induction. First, we define the Thue-Morse sequence as the result of repeated operations of $f(T)$, starting with $T_0 = 0$ (as in problem 1, part d). T_0 contains no overlapping factors. Now, for the induction hypothesis, assume that after $n-1$ iterations of $f(T)$, the sequence T_{n-1} has no overlapping factors. By the contrapositive of part (ii), $f(T_{n-1}) = T_n$ also has no overlapping factors. Hence, by induction we have that the Thue-Morse sequence as a whole has no overlapping factors. Finally, we showed in (b) that the presence of a cube in a sequence implies the presence of an overlapping factor, so the Thue-Morse sequence cannot contain any cubes either.

Miscellaneous

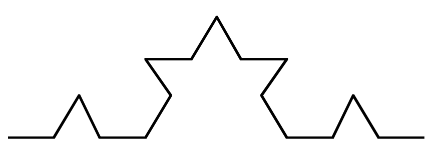
Just for fun, here are a few more cute and unexpected things you can do with the Thue-Morse sequence.

6. (a) The Koch snowflake is a well-known fractal that is constructed over iterations as follows. Our initial "snowflake", the zeroth iteration, is just a straight line segment.

In the first iteration, we take the middle third of the line segment, draw an equilateral triangle using that middle third as a base, and then erase the middle third, resulting in the following figure.



In the second iteration, we take every line segment in the above figure and repeat the same procedure: replacing the middle third of the line segment with the other two sides of the outward-facing equilateral triangle that has that middle third as a base.



In general, we create the nth iteration of the Koch snowflake by taking each line segment in the $(n-1)$th iteration and replacing the middle third by a "corner" in the shape of an equilateral triangle, in the same way as before.

(i) **[2]** Draw (no justification required) the third iteration of the Koch snowflake.

(ii) **[7]** A turtle reads the Thue-Morse sequence $t_0, t_1, \ldots$ and decides to crawl according to the sequence, as follows. At the nth step, if $t_n = 0$, it will crawl forward one unit and then turn 60 degrees to the left. If instead $t_n = 1$, it will turn 180 degrees (without moving). Prove that after 2^{2n+1} steps (that is, after following the sequence from $t_0, t_1, \ldots$ up to $t_{2^{2n+1}-1}$), the turtle will have traced out the nth iteration of the Koch snowflake. (Of course, we are ignoring the scale of the resulting snowflake here; we are only interested in its shape.)

(b) **[9]** Let $N = 2^{n+1}$. Let A_N be the set of integers i in $\{0, 1, \ldots, N-1\}$ such that $t_i = 0$, and let B_N be the set of integers j in $\{0, 1, \ldots, N-1\}$ such that $t_j = 1$. Prove that

$$\sum_{i \in A_N} i^k = \sum_{j \in B_N} j^k$$

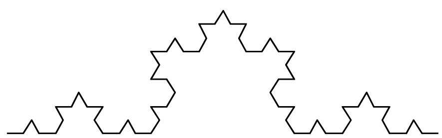
for all integers k from 1 to n. (This is a special case of the *Prouhet-Tarry-Escott problem.*)

(c) **[11]** As in the discussion after Problem 4, let $\{t'_n\}$ be the Thue-Morse sequence using $-1, 1$ instead of $0, 1$. Prove that

$$\left(\frac{1}{2}\right)^{t'_0}\left(\frac{3}{4}\right)^{t'_1}\left(\frac{5}{6}\right)^{t'_2}\cdots = \prod_{n=0}^{\infty}\left(\frac{2n+1}{2n+2}\right)^{t'_n} = \sqrt{2}.$$

Solution to Problem 6:

(a) (i)



(ii) We proceed by induction. The base case $n = 0$ is clear. Suppose that the turtle following the finite sequence $T_n = t_0 \cdots t_{2^{2n+1}-1}$ traces out the nth iteration of the Koch snowflake. Recall from Problem 1, part d, that $t_{2i}t_{2i+1}$ is always either 01 or 10, and that the sequence $T_a = t_0 t_1 \cdots t_{2^a-1}$ is generated from the sequence $T_{a-1} = t_0 t_1 \cdots t_{2^{a-1}-1}$ by simultaneously replacing each 0 with a 01 and each 1 with a 10. Then $T_{n+1} = t_0 t_1 \cdots t_{2^{2n+3}-1}$ is obtained from T_n by two iterations of this process.

Hence each pair $t_{2i}t_{2i+1}$ of consecutive instructions in T_n is replaced by eight consecutive instructions in T_{n+1} as follows: the pair 01 (step, turn 120° to the right) is replaced by 01101001 (step, turn 60° to the left, step, turn 120° to the right, step, turn 60° to the left, step, turn 120° to the right) and the pair 10 (turn 180°, step, turn 60° to the left) is replaced by 10010110 (turn 180°, step, turn 60° to the left, step, turn 120° to the right, step, turn 60° to the left, step, turn 60° to the left).

This replacement does not affect the orientation of this section of the path with respect to the rest of the path, since the turtle is oriented the same way at the beginning and the end, but it turns a single step—a single straight segment—into a sequence of four segments in the middle-third-replaced-by-corner shape. That is, the path the turtle traces following T_{n+1} is the same as the path the turtle traces following T_n, except that each segment in T_n is replaced by a sequence of four segments in the form of the Koch iteration. Since this is exactly how we get the $(n+1)$th iteration of the Koch snowflake from the nth iteration, this completes the induction.

(Note: an alternative solution is to induct by treating the $(n+1)$th iteration as a sequence of four copies of the nth iteration. In this case, one may check using the definition from Problem 1, part a that T_{n+1} also splits into four segments which produce four copies of the result of tracing T_n and join appropriately.)

(b) As in the discussion after Problem 4, let $\{t'_i\}$ be the Thue-Morse sequence using $-1, 1$ instead of $0, 1$. Then we wish to prove that

$$\sum_{i=0}^{N-1} t'_i i^k = 0$$

for all integers k from 1 to n. In fact, this is also true for $k=0$, as shown in Problem 5, part b(i). We proceed by induction on n. The base case $n=1$ is easy to check. Suppose the desired identity is true for $n-1$ (so $N=2^n$, $k=1,\dots,n-1$). Then for $N=2^{n+1}$ and any $k\in\{1,2,\dots,n\}$ we compute

$$\sum_{i=0}^{N-1} t'_i i^k = \sum_{i=0}^{2^n-1}(t'_i i^k + t'_{i+2^n}(i+2^n)^k) = \sum_{i=0}^{2^n-1}(t'_i i^k - t'_i(i+2^n)^k)$$

since $t'_{i+2^n} = -t'_i$ for $i=0,1,\dots,2^n-1$ by our first definition of the Thue-Morse sequence, and this is

$$\sum_{i=0}^{2^n-1} t'_i(i^k-(i+2^n)^k) = \sum_{i=0}^{2^n-1} t'_i(i-(i+2^n))(i^{k-1}+i^{k-2}(i+2^n)+\cdots+i(i+2^n)^{k-2}+(i+2^n)^{k-1})$$

$$= -2^n\sum_{i=0}^{2^n-1} t'_i P_{n,k}(i)$$

where $P_{n,k}(x) = x^{k-1}+x^{k-2}(x+2^n)+\cdots+x(x+2^n)^{k-2}+(x+2^n)^{k-1}$ is a polynomial of degree $k-1$. Write $P_{n,k}(x) = a_{k-1}x^{k-1}+a_{k-2}x^{k-2}+\cdots+a_1x+a_0$. Then the above summation is

$$\sum_{i=0}^{2^n-1} t'_i(a_{k-1}i^{k-1}+a_{k-2}i^{k-2}+\cdots+a_1 i+a_0) = a_{k-1}\sum_{i=0}^{2^n-1} t'_i i^{k-1}+\cdots+a_0\sum_{i=0}^{2^n-1} t'_i = 0$$

by the inductive hypothesis together with the known case $k=0$. This completes the induction.

(c) Write

$$P = \prod_{n=0}^{\infty}\left(\frac{2n+1}{2n+2}\right)^{t'_n}$$

and

$$Q = \prod_{n=1}^{\infty}\left(\frac{2n}{2n+1}\right)^{t'_n}.$$

Ignoring convergence issues which we will address later, we compute

$$PQ = 2\prod_{n=1}^{\infty}\left(\frac{n}{n+1}\right)^{t'_n} = 2\prod_{k=1}^{\infty}\left(\frac{2k}{2k+1}\right)^{t'_{2k}}\prod_{k=0}^{\infty}\left(\frac{2k+1}{2k+2}\right)^{t'_{2k+1}}$$

$$= 2\prod_{k=1}^{\infty}\left(\frac{2k}{2k+1}\right)^{t'_k}\prod_{k=0}^{\infty}\left(\frac{2k+1}{2k+2}\right)^{-t'_k} = 2\cdot\frac{Q}{P}.$$

Cancelling Q and solving for P gives $P^2=2$, or $P=\sqrt{2}$.

To make sure these manipulations are legitimate, we claim that if we rewrite P in the form

$$P = \prod_{k=0}^{\infty}\left(\frac{4k+1}{4k+2}\right)^{t'_{2k}}\left(\frac{4k+3}{4k+4}\right)^{t'_{2k+1}} = \prod_{k=0}^{\infty}\left(\frac{(4k+1)(4k+4)}{(4k+2)(4k+3)}\right)^{t'_{2k}}$$

then this two-by-two product is absolutely convergent. This is just because

$$\frac{(4k+1)(4k+4)}{(4k+2)(4k+3)} = \frac{16k^2+20k+4}{16k^2+20k+6} = 1 - \frac{2}{16k^2+20k+6} = 1 - \frac{1}{8k^2+10k+3}$$

and similarly

$$\frac{(4k+2)(4k+3)}{(4k+1)(4k+4)} = 1 + \frac{1}{8k^2+10k+2}$$

and we know that

$$\prod_{k=0}^{\infty}\left(1 \pm \frac{1}{8k^2+10k+5/2 \mp 1/2}\right)$$

is absolutely convergent if and only if

$$\sum_{k=0}^{\infty}\left(\pm\frac{1}{8k^2+10k+5/2 \mp 1/2}\right)$$

is absolutely convergent, which we can see is true since $\sum_{n=1}^{\infty}\frac{1}{n^2}$ is absolutely convergent. Essentially the same argument gives that Q is absolutely convergent upon being written as a product of pairs of consecutive terms as well. Since the manipulations we performed above can also be written as rearrangements of pairs of consecutive terms, this is acceptable.