Survey of combinatorics on words and patterns

Menghua Wu

May 17, 2018

Abstract

We survey recent and historical results from combinatorics on words, with a focus on patterns in infinite words. The works of Axel Thue in the early twentieth century gave rise a plethora of research around avoidable and unavoidable patterns in infinite words, as well as bounds on and characterizations of such patterns. For example, Zimin proved in 1984 that all unavoidable patterns must be contained within a Zimin word, also known as a sesquipower. Finally, we conclude with an overview of applications of words and patterns to other branches of mathematics.

1 Introduction

The study of combinatorics on words is a field of mathematics that focuses on words, patterns, and formal languages. In this paper, we provide a survey of results related to patterns in infinite words over finite alphabets.

Formally, a word is a sequence of symbols composed from a finite set Σ, and a factor of word w is any sequence of consecutive symbols in w [4]. We denote Σ^* as the set of all words over Σ, or $\Sigma^* = \epsilon \cup \left(\cup_k \Sigma^k\right), \forall k \geq 1$, where ϵ is the empty string. A pattern p is a word over Σ with the following properties. We say a word w contains p if there is a mapping from each symbol in p to a non-empty word, such that the concatenated words form a factor of w. On the contrary, a word w avoids pattern p if it does not contain p [6]. For example, the word banana contains the pattern xx but avoids the pattern xxx: both $\mathsf{x} \mapsto \mathsf{na}$ or $\mathsf{x} \mapsto \mathsf{an}$ are valid mappings. The mapping from pattern to word is commonly known as a substitution or a morphism. A substitution h obeys the properties that $h(xy) = h(x)h(y)$ and if $x = a_0a_1, \ldots$, then $h(x) = h(a_0)h(a_1)\ldots$ [4]. We may proceed with these basic concepts in mind.

The paper is organized as follows. In section 2, we introduce combinatorics on words through early studies of square-free words. In section 3, we continue our study of patterns and their avoidability through Zimin words and bounds on words containing the n^{th} Zimin word. We conclude in section 4 by providing applications of one pattern in particular, the Thue-Morse word.

2 Square-free words

Combinatorics on words stemmed from the works of Axel Thue in 1906 and 1912 [12, 13], which we study through modern translations [3, 10]. In his 1906 paper, Thue famously proposed an infinite word over an alphabet of four symbols that avoids the pattern xx. Words that exhibit this property are known as square-free words.

It is simple to show that all infinite words over alphabets of one and two symbols must contain the pattern xx. The one symbol case is trivial. Now suppose $\Sigma = \{0, 1\}$. Then every block of four symbols must contain xx, with representative examples depicted below.

$$\begin{array}{cccc} 0000 & 0101 & 0110 & 0111 \\ \mathsf{x} \mapsto 0 & \mathsf{x} \mapsto 01 & \mathsf{x} \mapsto 1 & \mathsf{x} \mapsto 1 \end{array}$$

Thus we might ask, what is the smallest number of symbols required to find an infinite word that avoids the pattern xx? Thue showed that three symbols suffice [12].

Theorem 2.1 (Thue, 1906)**.** *An arbitrary long square-free exists over three symbols.*

We prove a stronger result.

Theorem 2.2 (Thue, 1906)**.** *An arbitrary long word exists over alphabet $\Sigma_3 = \{a, b, c\}$ that avoids patterns aca and bcb.*

Proof. We provide a constructive proof. Let $u \in \Sigma_3^*$ be a square-free word that avoids patterns aca and bcb.

1. Replace each instance of ac with $\mathsf{a}\beta\alpha$, and replace each instance of bc with $\mathsf{b}\alpha\beta$, to construct u'. For example, if $u = \mathsf{abc}$, then $u' = \mathsf{ab}\alpha\beta$. Note that we can reconstruct u from u' by removing all α's and replacing all β's with c, or removing β's and replacing α's.

   We claim that u' is still square-free and avoids patterns $\mathsf{x}\alpha\mathsf{x}$ and $\mathsf{x}\beta\mathsf{x}$. If u' contained the pattern xx, then the reconstructed u would contain a square, which is a contradiction. Without loss of generality, suppose u' contained $\mathsf{x}\alpha\mathsf{x}$. The α must be preceded or succeeded by an β, so $\mathsf{x}\alpha\mathsf{x} = \mathsf{y}\beta\alpha\mathsf{y}\beta$ or $\mathsf{x}\alpha\mathsf{x} = \beta\mathsf{y}\alpha\beta\mathsf{y}$. When we delete all α's and replace β's with c's, we obtain a factor of ycyc in w, which we assumed to be square-free. Thus our claim holds.

2. Insert γ after every symbol in u' to construct u''. For example, if $u' = \mathsf{ab}\alpha\beta$, then $u'' = \mathsf{a}\gamma\mathsf{b}\gamma\alpha\gamma\beta\gamma$. u'' is obviously square-free; otherwise we could delete the γ's to reconstruct u', which would contain a square.

3. Replace each a in u'' with $\alpha\beta\alpha$ and replace each b with $\beta\alpha\beta$ to construct w. For example, if $u'' = \mathsf{a}\gamma\mathsf{b}\gamma\alpha\gamma\beta\gamma$, then $w = \alpha\beta\alpha\gamma\beta\alpha\beta\gamma\alpha\gamma\beta\gamma$.

   We claim that w is square-free and avoids factors $\alpha\gamma\alpha$ and $\beta\gamma\beta$.

   First we prove the latter claim. Observe that in u', symbols $\{\mathsf{a}, \alpha\}$ alternate with symbols $\{\mathsf{b}, \beta\}$, else we would have a square in u'. After we insert a γ between every two symbols in u'', all three-symbol factors take the form $\mathsf{a}\gamma\mathsf{b}, \mathsf{a}\gamma\beta, \alpha\gamma\mathsf{b}, \alpha\gamma\beta$ or their reversals. Thus, the only factors in w containing γ are $\alpha\gamma\beta$ and $\beta\gamma\alpha$.

   Now we show that w is square-free. Assume for contradiction that w contains pattern xx. Observe that between any two γ in w, there are only four possible factors: $\alpha, \beta, \alpha\beta\alpha$, and $\beta\alpha\beta$, where the latter two follow from a and b, respectively. Since none of these four factors contain a square, x must contain at least one γ.

   Without loss of generality, suppose x only contained a single γ, succeeded by an α (the same argument holds if γ is preceded by a symbol, and if we take β instead). Then xx must contain either $\gamma\alpha\gamma\alpha$ or $\gamma\alpha\beta\alpha\gamma\alpha$, and w must contain $\alpha\gamma\alpha$, which we have established to be false. Thus xx must contain at least two occurrences of γ. With $X = \{\alpha, \beta, \alpha\beta\alpha, \beta\alpha\beta\}$,

$$\mathsf{x} = \mathsf{p}\gamma x_1\gamma \ldots \gamma x_m\gamma\mathsf{q}$$

for $m \geq 1$, where $x_1, \ldots, x_m \in X$, $\mathsf{qp} \in X$ (between two copies of x), and $\mathsf{p'p}, \mathsf{qq'} \in X$ for some $\mathsf{p'}, \mathsf{q'}$ (boundaries of xx).

If $\mathsf{q} = \epsilon$, then $\mathsf{p}\gamma x_1 \gamma \ldots \gamma x_m$ causes a square in u'', and likewise for $\mathsf{p} = \epsilon$. So $\mathsf{qp} = \alpha\beta\alpha$ or $\mathsf{qp} = \beta\alpha\beta$. Without loss of generality, suppose $(\mathsf{q}, \mathsf{p}) = (\alpha, \beta\alpha)$. By our construction, w cannot start with $\mathsf{p}\gamma = \beta\alpha\gamma$, so w contains the factor $\mathsf{qp}\gamma x_1 \gamma \ldots \gamma x_m \gamma$. In this case, u'' again contains a square, so we reach a contradiction, and w must be square-free. Thus we have shown that given any square-free word $u \in \Sigma_3^*$ avoiding aca and bcb, we can construct a longer square-free word $w \in \Sigma_3^*$ with the same properties.

□

Theorem 2.3 (Thue, 1906)**.** *An infinite square-free word exists over three symbols. That is, there exists a sequence $\{w_n\}_{n \geq 0}$ such that w_n is a prefix of w_{n+1} and $\forall i \geq 0$, w_i is square-free.*

Proof. Let u be a square-free word over the alphabet $\{\mathsf{a}, \mathsf{b}, \mathsf{c}\}$ that avoids aca, bcb, and starts with symbols a or b. We construct a new word $\sigma(u)$ with the morphism σ, where

$$\begin{aligned} \sigma(\mathsf{a}) &= \mathsf{abac} \\ \sigma(\mathsf{b}) &= \mathsf{babc} \\ \sigma(\mathsf{c}) &= \mathsf{bcac} \quad \text{if } \mathsf{c} \text{ is preceded by } \mathsf{a} \\ \sigma(\mathsf{c}) &= \mathsf{acbc} \quad \text{if } \mathsf{c} \text{ is preceded by } \mathsf{b}. \end{aligned}$$

For example, if we take $u = \mathsf{abc}$ as in the previous proof, we obtain $\sigma(u) = \mathsf{abac\ babc\ acbc}$.

Upon inspection, we see that $\sigma(u)$ is the same as the w created by the proof of theorem 2.2, where α, β, γ are replaced by a, b, c, respectively. Thus, σ creates words that are square-free. If we let $w_0 = \mathsf{a}$, then we have an infinite sequence of square-free words, $w_{n+1} = \sigma^n(w_0)$, where w_n is the prefix of $w_{n+1}, \forall n \geq 0$.

□

Thue's 1912 paper introduced an alternative construction of an infinite square-free word, as well as the sequence now known as Thue-Morse word [3].

Definition 2.1 (Thue-Morse word)**.** *Let $\{\mathsf{a}, \mathsf{b}\}$ be a two symbol alphabet. Consider the substitution h*

$$h(\mathsf{a}) = \mathsf{b} \qquad h(\mathsf{b}) = \mathsf{a}.$$

For $n \geq 0$, let $u_n = h^n(\mathsf{a})$ and $v_n = h^n(\mathsf{b})$, such that

$$\begin{aligned} u_0 &= \mathsf{a} & v_0 &= \mathsf{b} \\ u_1 &= \mathsf{ab} & v_1 &= \mathsf{ba} \\ u_2 &= \mathsf{abba} & v_2 &= \mathsf{baab} \\ &\cdots \end{aligned}$$

and the general terms are

$$u_{n+1} = u_n v_n \qquad v_{n+1} = v_n u_n.$$

Thue proved that the word u_n is overlap-free—that is, u_n avoids the pattern xyxyx for $\mathsf{x} \neq \epsilon$. Furthermore, Thue derived a relationship between overlap-free words and square-free words.

Theorem 2.4 (Thue, 1912)**.** *Let w be an infinite overlap-free word. Then the reverse image of w under the substitution g*

$$g(\mathsf{a}) = \mathsf{abb} \quad g(\mathsf{b}) = \mathsf{ab} \quad g(\mathsf{c}) = \mathsf{a}$$

is an infinite square free word on three symbols.

That is, $g(w') = w$, where w' is square-free and w is overlap-free. Let us consider this theorem in context of the Thue-Morse word, which extends infinitely as

$$w = \mathsf{abba\ baab\ baab\ abba\ baab\ abba\ abba\ baab} \ldots$$

From w we may apply the "reverse substitution" as

$$\underbrace{\mathsf{abb}}_{\mathsf{a}}\underbrace{\mathsf{ab}}_{\mathsf{b}}\underbrace{\mathsf{a}}_{\mathsf{c}}\underbrace{\mathsf{abb}}_{\mathsf{a}}\underbrace{\mathsf{a}}_{\mathsf{c}}\underbrace{\mathsf{ab}}_{\mathsf{b}}\ldots$$

to obtain square-free word $w' = \mathsf{abcacbabcbac}. \ldots$

3 Unavoidable patterns

Thue's work in the early twentieth century was forgotten for some decades, with partial revivals in the 1920s and 1940s, and renewed interest since the 1980s [4]. Much of the work we present here originates from this most recent period. Unavoidable patterns were characterized independently by Bean, Ehrenfeucht and McNulty [2] and Zimin [14]. For convenience, the rest of this section is based upon the latter's work.

We give an overview of Zimin's characterization of unavoidable patterns, provide an algorithm for determining unavoidability, and summarize recent bounds on the length of words containing Zimin words.

3.1 Zimin words

To generalize the idea of square-free words, we say a pattern p is q-avoidable if every sufficiently long word over q symbols avoids p, for $q \in \mathbb{N}$. Conversely, a pattern is q-unavoidable if all long words over q symbols contains p. A pattern is unavoidable if it is q-unavoidable for all q. For example, we showed that the pattern xx is 2-unavoidable but 3-avoidable. Thus, xx is not unavoidable.

Definition 3.1. *A Zimin word, also known as a sesquipower, is defined recursively as*

$$\begin{aligned} Z_1 &= x_1 \\ Z_2 &= x_1 x_2 x_1 \\ &\ldots \\ Z_n &= Z_{n-1} x_n Z_{n-1} \end{aligned}$$

over an alphabet $\Sigma = \{x_1, \ldots, x_N\}$.

Proposition 3.1 (Zimin, 1982)**.** *All Zimin words are unavoidable.*

Proof. We present a simple inductive proof from [9].

Claim 3.1. *The pattern $Z_2 = \mathsf{xyx}$ is unavoidable.*

Proof of claim. Consider an arbitrary alphabet Σ on k symbols. Let $w \in \Sigma^*$ be a word of length $2k + 1$. Then some symbol a must be repeated three times, and we can factor w as

$$w = w_1 \mathsf{a} w_2 \mathsf{a} w_3 \mathsf{a} w_4.$$

The pattern xyx appears with the substitution h,

$$h(\mathsf{x}) = \mathsf{a} \qquad h(\mathsf{y}) = w_2 \mathsf{a} w_3.$$

□

Claim 3.2. *If* p *is unavoidable on alphabet* Σ *and* α *is a symbol that does not occur in* p*, then* pαp *is unavoidable on* Σ*.*

Proof of claim. Since p is unavoidable on Σ, there is an integer ℓ such that all words $u \in \Sigma^*$ of length ℓ contain p. Now consider the word

$$w = w_1 s_1 w_2 s_2 \dots s_{k^\ell} w_{k^\ell+1}$$

where $|w_i| = k$ and s_i is a single symbol, $\forall i \in \left\{1, 2, \dots, k^\ell + 1\right\}$. There are at most k^ℓ unique words of length ℓ, so there must exist distinct i, j such that $w_i = w_j$. Since w_i and w_j are separated by at least one symbol, w contains pattern xyx. □

We can seen that $Z_1 = \mathsf{x}$ is trivially unavoidable, and claim 3.1 showed that $Z_2 = \mathsf{xyx}$ is unavoidable. We may apply claim 3.2 inductively to show that Z_n is unavoidable over all alphabets.

□

3.2 Zimin algorithm

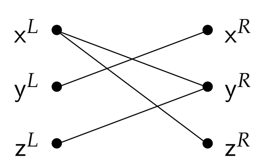
In terms of computability, it turns out that avoidability is decidable through a property we denote as irreducibility. Besides his eponymous Zimin words, Zimin developed a recursive algorithm for determining whether a pattern is reducible, and thereby avoidable.

Theorem 3.1 (Zimin, 1982)**.** *A pattern is avoidable if and only if it is irreducible.*

We provide reduction algorithm below. Let $p \in \Sigma^*$ be a pattern. We construct the adjacency graph $G(p)$ as follows.

1. For each symbol $\mathsf{x} \in \Sigma$, add two vertices, x^L and x^R. Denote the left and right sets of vertices as Σ^L and Σ^R, respectively.
2. For symbols $\mathsf{x}, \mathsf{y} \in \Sigma$, add an edge between x^L and y^R if and only if xy is a factor of p.

For example, the adjacency graph of xyxzyx is depicted below.



We denote a nonempty subset $F \subseteq \Sigma$ as a free set if for all symbols $\mathsf{x}, \mathsf{y} \in F$, there is no path connecting x^L to y^R. In the above example, we have free sets $\{\mathsf{x}\}$ and $\{\mathsf{y}\}$.

Given a pattern p and a free set F, we can create a new pattern q by deleting all symbols in F from p. Here, we say p reduces in one-step to q. If there is a sequence of one-step reductions from p to q, when we say p reduces to q. In context of this proof, a pattern p is reducible if p reduces to ϵ.

Note, however, that we only require one such sequence of reductions to exist. For example, consider our previous example p = xyxzyx. If we first delete $\{\mathsf{b}\}$, we would obtain an irreducible pattern q = xxzx, but if we delete $\{\mathsf{a}\}$ first, we reduce our pattern to yzy, which in turn reduces to ϵ. Therefore, we require a recursive algorithm to check all paths.

3.3 Tower-type bounds

While an algorithm suffices to check whether a pattern is avoidable, it is often more convenient to cite Zimin's following theorem in proving bounds on avoidability.

Theorem 3.2 (Zimin, 1982)**.** *A pattern is unavoidable if and only if it is contained within a Zimin word.*

Thus we might consider quantitative bounds on patterns containing Zimin words. For a fixed word Z_n, let $f(n,q)$ be the smallest number such that every word of $f(n,q)$ over q symbols contains Z_n [7].

Theorem 3.3 (Cooper and Rorabaugh, 2014)**.** *For any $n \geq 3$ and $q \geq 2$,*

$$f(n,q) \leq \underbrace{q^{q^{\cdot^{\cdot^{\cdot^{2q+o(q)}}}}}}_{(n-1)}$$

where the $o(q)$ term is independent of n and $\lim_{q\to\infty} o(q) \to 0$.

We first prove the following lemma.

Lemma 3.1. $f(n+1,q) \leq (f(n,q)+1)\left(q^{f(n,q)}+1\right) - 1.$

Proof of lemma. Consider the word

$$w = \underbrace{\underbrace{\mathsf{ab}\ldots\mathsf{c}}_{f(n,q)}\mathsf{x}\underbrace{\mathsf{de}\ldots\mathsf{f}}_{f(n,q)}\mathsf{y}\ldots\mathsf{z}\underbrace{\mathsf{rs}\ldots\mathsf{t}}_{f(n,q)}}_{q^{f(n,q)+1}}$$

composed of $q^{f(n,q)+1}$ words, each of length $f(n,q)$, separated by a single symbol. It is easy to see that w has length $(f(n,q)+1)\left(q^{f(n,q)+1}\right) - 1$. Each of the smaller words must contain Z_n, by definition, and there are $q^{f(n,q)}$ unique words of length $f(n,q)$ over q symbols. So by the pigeonhole principle, w must contain two copies of the same word, separated by at least one symbol. Therefore, w must contain Z_{n+1}.

□

Proof of theorem. We prove the theorem by induction. The base case is $f(2,q) = 2q+1$, in which pattern xyx appears as two blocks of q symbols, separated by a single symbol.

We apply lemma 3.1 recursively to attain the desired bound,

$$f(n,q) \leq \underbrace{q^{q^{\cdot^{\cdot^{\cdot^{2q+o(q)}}}}}}_{(n-1).}$$

□

Following Cooper and Rorabaugh, there have been several efforts to provide tighter bounds on $f(n,q)$. In particular, we highlight the contributions of Rytter and Shur [11] and Conlon, Fox, and Sudakov [6].

Lemma 3.2 (Rytter and Shur, 2015)**.** $f(3,q) \leq 2^{q+1}(q+1)!$

Proof. We say that a word w is 2-minimal if it contains Z_2, but all of its factors avoid Z_2. So w must necessarily be of the form aaa, for a single symbol a, or $\mathsf{ab}_1^{j_1}\ldots\mathsf{b}_r^{j_r}\mathsf{a}$, where the b_i are distinct and $j_i \in \{1,2\}, \forall i$.

Let $t(2,q)$ be the number of 2-minimal words over an alphabet of q symbols. There are q words of the first form and for a given r, there are $q \cdot (q-1)(q-2)\ldots(q-r)\cdot 2^r$ words of the second form. So

$$t(2,q) = q + \sum_{r=1}^{q-1} q(q-1)\ldots(q-r)2^r \le q! \sum_{r=0}^{q-1} 2^r \le 2^q q! - 1.$$

Now consider the word

$$w = \underbrace{\underbrace{\mathsf{ab}\ldots\mathsf{c}}_{f(2,q)}\mathsf{x}\underbrace{\mathsf{de}\ldots\mathsf{f}}_{f(2,q)}\mathsf{y}\ldots\mathsf{z}\underbrace{\mathsf{rs}\ldots\mathsf{t}}_{f(2,q)}}_{t(2,q)+1}$$

where each word of length $f(2,q)$ must contain at least one 2-minimal word. By the pigeonhole principle, two of these $t(2,q)+1$ words must be the same. So we find a copy of $Z_3 = \mathsf{xyx}$ where the repeated words constitute the x and symbols between them form the y.

With some simple algebra, we have

$$f(3,q) \le (f(2,q)+1)(t(2,q)+1) - 1 \le (2q+2)2^q q! = 2^{q+1}(q+1)!$$

as desired.

□

Theorem 3.4 (Conlon, Fox, and Sudakov, 2017)**.** *For any $n \ge 3, q \ge 35$,*

$$f(n,q) \le \underbrace{q^{q^{\cdot^{\cdot^{\cdot^{q-o(q)}}}}}}_{(n-1)}.$$

Sketch of proof. We prove by induction that

$$qf(n,q) \le \underbrace{q^{q^{\cdot^{\cdot^{\cdot^{q}}}}}}_{(n-1)}.$$

from which the result follows.

The base case is $n = 3$, which directly follows from lemma 3.2:

$$f(3,q) \le 2^{q+1}(q+1)! \le q^{q-1}$$

when $q \ge 35$. Now suppose there is some $T \ge q^q$ for which $qf \le T$. By lemma 3.1,

$$f(n+1,q) \le (f+1)(q^f+1) - 1 = fq^f + q^f + f \le q^{T-1}$$

where $f = f(n,q)$ for concision. Plugging in T, we may write

$$f(n+1,q) \le (T/q+2)q^{T/q} \le Tq^{T/q} \le q^{T-1}.$$

Our result follows from $f(n+1,q) \le q^{T-1}$.

□

While these results provide bounds on the maximum length of a word before Z_n appears, the longest word is not indicative of most words. Instead, [6] uses a birthday paradox argument to determine the threshold length that Z_n appears in a random word.

Theorem 3.5 (Conlon, Fox, and Sudakov, 2017)**.** *For fixed n and q tending to infinity, the threshold length for Z_n to appear in a random word over q symbols is $q^{2^{n-1}-(n+1)/2}$.*

Proof. Recall that $Z_n = Z_{n-1}x_nZ_{n-1}$, where two copies of Z_{n-1} are separated by a symbol, and each symbol x_i is mapped to some word. The minimal Z_{n-1} maps each x_i to one symbol, so the length of such a Z_{n-1} is $2^{n-1}-1$. There are $D = q^{n-1}$ distinct copies of such Z_{n-1}, so the probability that any word of length $2^{n-1}-1$ contains Z_{n-1} is

$$p = \frac{q^{n-1}}{q^{2^{n-1}-1}}.$$

For a random word of length N, where $N \geq 2^{n-1}-1$, we expect to find Np copies of Z_{n-1}.

In the birthday paradox, it becomes probable that we encounter two people of the same birthday when there are approximately $D^{1/2}$ people, where D is the number of days in a year. Likewise, for $D = q^{n-1}$, it becomes probable we encounter two copies of Z_{n-1} when $Np = D^{1/2}$. Solving for N, we obtain

$$N = q^{2^{n-1}-(n+1)/2}.$$

□

In context, consider copies of Z_3 in the English alphabet with $q = 26$. According to theorem 3.3, it is only guaranteed that we find Z_3 with at least 10^{34} characters, but by theorem 3.5, it is likely that Z_3 will appear within 1000 characters.

4 Thue-Morse word

Among all patterns encountered in combinatorics on words, none is more well-represented in other fields of mathematics than the Thue-Morse, whose definition we gave in section 2. This section introduces a few appearances of the Thue-Morse word throughout mathematics, but it is by no means comprehensive. There are entire papers written on this subject [1], and we have selected a few interesting cases for study.

4.1 Infinite play in chess

The Thue-Morse word appears in many disparate fields of mathematics from algebra to number theory—and even chess [1]. In an obscure 1929 paper, chess grandmaster Max Euwe independently discovered the Thue-Morse word to show that infinite play may exist in chess under some reasonable conditions [8].

A supposed German rule dictated that a draw in chess would occur if any sequence of moves repeated three times in a row. After all, people assumed that if a game continued for long enough, then some sequence would eventually repeat. Euwe proved that this was not the case.

Theorem 4.1 (Euwe, 1929)**.** *An infinite, cube-free sequence of chess moves exists.*

Proof. Suppose we encoded a chess game as a word over two symbols, such that each symbol represented a sequence of chess moves. For example, let

$$\begin{aligned}\mathsf{a} &: A1 - B2, B2 - A1\\ \mathsf{b} &: C1 - D2, D2 - C1\end{aligned}$$

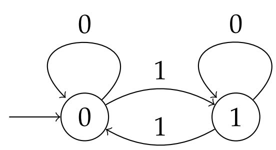
be a sequence where bishops move back and forth. Now consider the Thue-Morse word on $\{\mathsf{a}, \mathsf{b}\}$. As we stated in section 2, this word is overlap-free, so it is trivially cube-free. Therefore, there exists an infinite, cube-free sequence of chess moves, and the aforementioned rule is invalid for enforcing a draw. $\square$

4.2 Finite automata

Consider the Thue-Morse word t over binary alphabet $\{0, 1\}$ and let t_n be the $(n+1)^{\text{th}}$ symbol from the left, so that $t = t_0 t_1 t_2 \ldots$. An alternate definition for the Thue-Morse word is

$$t_n = \#\ \text{1s in binary expansion of } n \mod 2.$$

An interesting result from this representation relates the Thue-Morse word to finite automata depicted below.



In computability theory, finite automata read in input symbols and change states depending on its transition function (edges) and the symbols read. For example, the word 101 would cause the above automaton to transition first to state 1, stay at 1, and transition to state 0.

We can easily see that the above finite automaton outputs the number of 1s in the input, modulo 2. So via the alternative definition for t_n, we find that t_n is the state reached when the input is the binary representation of n. Sequences that may be output by a finite automaton are known as automatic. Works by Cobham showed that sequences are automatic if and only if they are the fixed points of morphisms; as we have seen above, the Thue-Morse word may be mapped to itself, so it satisfies this property from both directions [5].

References

[1] Jean-Paul Allouche, , and Jeffrey Shallit, *The ubiquitous prouhet-thue-morse sequence*, (1999), 1–16.

[2] Dwight R. Bean, Andrzej Ehrenfeucht, and George F. McNulty, *Avoidable patterns in strings of symbols.*, Pacific J. Math. **85** (1979), no. 2, 261–294.

[3] Jean Berstel, *Axel thue's papers on repetitions in words: a translation*, Publications du LaCIM, Universite du Quebec a Montreal (1995).

[4] Jean Berstel and Dominique Perrin, *The origins of combinatorics on words*, European Journal of Combinatorics **28** (2007), no. 3, 996 – 1022.

[5] Alan Cobham, *Uniform tag sequences*, Mathematical systems theory **6** (1972), no. 1, 164–192.

[6] D. Conlon, J. Fox, and B. Sudakov, *Tower-type bounds for unavoidable patterns in words*, (2017).

[7] J. Cooper and D. Rorabaugh, *Bounds on Zimin Word Avoidance*, ArXiv e-prints (2014).

[8] Machgielis Euwe, *Mengentheoretische betrachtungen uber das schachspiel*, Proc. Konin. Akad. Wetenschappen (1929), 633–642.

[9] M. Lothaire, *Algebraic combinatorics on words*, Encyclopedia of Mathematics and its Applications, Cambridge University Press, 2002.

[10] James F. Power, *Thue's 1914 paper: a translation*, CoRR **abs/1308.5858** (2013).

[11] Wojciech Rytter and Arseny M. Shur, *On searching zimin patterns*, CoRR **abs/1409.8235** (2014).

[12] Axel Thue, *Uber unendliche zeichenreihen*, Christiana Videnskabs Selskabs Skrifter (1906), no. 7.

[13] ———, *Uber die gegenseitige lage gleicher teile gewisser zeichenreihen*, Christiana Videnskabs Selskabs Skrifter (1912), no. 1.

[14] A. I. Zimin, *Blocking sets of terms*, Mat. Sb. **119(161)** (1984), 363–375, Translated in *Sb. Math*. vol. 47 (1984), 353-364.