For us a sequence is (usually) a map from $\mathbb{N} = \{0, 1, 2, \ldots\}$ to $\Sigma$, where $\Sigma$ is a (usually finite) alphabet. (Sometimes we index starting at 1 instead of 0; sometimes we talk about "two-sided" or "bi-infinite" sequences, which are maps from $\mathbb{Z}$ to $\Sigma".")

A sequence is also called an infinite word (infinite string).

**Example:** The characteristic sequence of the prime numbers $P = \{2, 3, 5, 7, 11, \ldots\}$:

<table>
<thead>
<tr>
<th>$n$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_n$</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

We write $P = (P_n)_{n \geq 0}$.

**Example:** The Thue-Morse sequence

$$(S_2(n) \mod 2)_{n \geq 0} = (t_n)_{n \geq 0}$$

Here $S_2(n)$ is the sum of the bits of $n$ expressed in base 2.
\[
\begin{align*}
  n &= 0 \ 1 \ 2 \ 3 \ 4 \ 5 \ 6 \ 7 \ 8 \ \cdots \\
  S_2(n) &= 0 \ 1 \ 1 \ 2 \ 1 \ 2 \ 2 \ 3 \ 1 \ \cdots \\
  t_n &= 0 \ 1 \ 1 \ 0 \ 1 \ 0 \ 0 \ 1 \ 1 \ \cdots 
\end{align*}
\]

Note: \[t_0 = 0\]
\[t_{2n} = t_n \quad n \geq 0 \quad (*)\]
\[t_{2n+1} = 1 - t_n\]

Informally, a sequence is \underline{k-automatic} (k \geq 2) if
(a) it has finite range
(b) it satisfies a system of equations like those in (*) where the left and right-hand sides involve subsequences with indices of the form \(2^i \cdot n + j\), \(0 \leq j < 2^i\).

k-automatic sequences have many, many interesting properties. Some of them are "pseudo-random", lying in between the very simplest sequences (the ultimately periodic ones) and the most complicated sequences (random sequences).
A sequence \((a_n)_{n \geq 0}\) is ultimately periodic if there exist integers \(p, N\) such that \(a_i = a_{i+p} \quad \forall i \geq N\).

Sometimes \(p\) is called the period and \(N\) the preperiod.

Confusingly, sometimes "period" refers to the word
\[
\ldots a_N \ a_{N+1} \ldots \ a_{N+p-1}
\]
and sometimes "preperiod" refers to
\[
\ldots a_0 \ a_1 \ldots a_{N-1}
\]

Example: The Rudin–Shapiro sequence \((r_n)_{n \geq 0}\)

is defined to be \((a_{11}(n) \mod 2)_{n \geq 0}\)

where \(a_{11}(n)\) is the number of occurrences of the block '11'

in the binary representation of \(n\). (Warning: sometimes

the Rudin–Shapiro sequence is defined to be

\[
r'_n = (-1)^{a_{11}(n)}
\]

in the literature, which is just a recoding of the sequence over the alphabet \((-1, 1\)\).)

\[
\begin{align*}
n & 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 \\
a_{11}(n) & 0 0 0 1 0 0 1 2 0 0 0 1 1 1 2 3 \\
r_n & 0 0 0 1 0 0 1 0 0 0 1 1 1 0 1
\end{align*}
\]

Occurrences may overlap, so '11' has 2 occurrences of '11'.
It is automatic because

\[ r_{2n} = r_n \]
\[ r_{4n+1} = r_n \quad \forall n \geq 0 \]
\[ r_{8n+3} = 1 - r_n \]
\[ r_{8n+7} = 1 - r_{4n+3} \]

\((r_n)_{n \geq 0}\) has an amazing pseudo-randomness property:

\[ \sum_{0 \leq n < N} \left[ r_n = r_{n+c} \right] = \frac{N}{2} + o(N) \]

for each integer \(c \geq 1\), where the \(o(N)\) in the little-

\(o\) can depend on \(c\). Here \(\left[ x = y \right]\) is the so-called

Iverson bracket, which evaluates to 1 if \(x = y\)

and 0 if \(x \neq y\).

On the next page here is a plot of

\[ \left( \sum_{0 \leq n < N} \left[ r_n = r_{n+1} \right] \right) - \frac{N}{2} \]

for \(N = 0, 1, \ldots, 2^{14} - 1\).

(See Høholdt et al., IEEE Trans. Info. Theory IT-31 (1985),
549-552.)
In this course we will study the computational, logical, algebraic, combinatorial, and number-theoretic properties of automatic sequences and their generalizations. Generalizations include the morphic sequences and the $k$-regular sequences.

**Computational:**

Automatic sequences get their name from another way to view such sequences: as computed by a DFAO (deterministic finite automaton with output). For example, here is a 2-DFAO computing the Rudin–Shapiro sequence:

![Diagram of a 2-DFAO computing the Rudin–Shapiro sequence]

A state label like "$q_i / a" means

- state name is $q_i$
- output is $a$.

To use this machine:
- express $n$ in base 2
- feed into DFAO starting with initial state $q_0$
- follow arrows
- output is output on last state reached.
The kinds of computational questions we are interested in include:
- given a sequence, is it $k$-automatic for some $k$?
- how many states are needed to represent a sequence, if it is representable?
- can a sequence be $k$-automatic and $k'$-automatic for $k \neq k'$?
- what is the blow-up in the number of states when converting from one representation of a $k$-automatic sequence to another?
- does it matter if numbers are read least-significant-digit first, instead of most significant digit first? ("lsd" versus "msd")
- if a sequence is not $k$-automatic, how can we prove this?
- what happens when we "transduce" an automatic sequence with a finite-state transducer?
The kinds of logical questions we are interested in include:

- Is there a logical characterization of k-automatic sequences?
- What logical theories involving automata are decidable/undecidable?

Ex. $\text{FO}(\mathbb{N}, +, V_2)$ is decidable but $\text{FO}(\mathbb{N}, +, V_2, V_3)$ is undecidable.

Here $V_k(n) = \max \{ k^i : k^i \mid n \}$.

- What questions about automatic sequences are decidable?

The kinds of algebraic questions we are interested in include the following:

- What are the relationships between automata, semigroups, and monoids?

- Let $(a_n)_{n \geq 0}$ be a sequence. Under what conditions is the formal power series \[ \sum_{n \geq 0} a_n x^n \] algebraic over $\text{GF}(q)[x]$? Here $q$ is a prime power and $\text{GF}(q)$ is the finite field of $q$ elements.

Here is an example. Let $q = 2$ and $a_n = t^n$, the Thue-Morse sequence.
Let \( T(x) = \sum_{n \geq 0} t_n x^n \).

Then

\[
T(x) = \sum_{n \geq 0} t_{2n} x^{2n} + \sum_{n \geq 0} t_{2n+1} x^{2n+1}
\]

\[
= \sum_{n \geq 0} t_n x^{2n} + x \sum_{n \geq 0} (1 - t_n) x^{2n}
\]

\[
= \left( \sum_{n \geq 0} t_n x^n \right)^2 + x \sum_{n \geq 0} x^{2n} + x \left( \sum_{n \geq 0} t_n x^n \right)^2
\]

\[
= T(x)^2 + \frac{x}{1 - x^2} + xT(x)^2
\]

\[\Rightarrow (1 + x) T(x)^2 + T(x) + \frac{x}{1 + x^2} = 0\]

\[\Rightarrow (1 + x)(1 + x^2) T(x^2) + (1 + x^2) T(x) + x = 0\]

So \( T \) is the root of a quadratic equation with coefficients in \( \text{GF}(2) \) [\( x \)].

We will see this is more generally true for all \( q \)-automatic sequences, where \( q \) is a prime power.
The kinds of combinatorial questions we are interested in include:

- What sort of repetitions can occur in an automatic sequence?

- How many different blocks of size n occur? (subword complexity)

- What is the frequency of occurrence of a letter or block in an automatic sequence?

- What are the palindromic that occur in an infinite sequence? The primitive words? etc.

- What are the blocks shared in common between two automatic sequences?

- Is a given sequence recurrent (i.e., every block that occurs, occurs infinitely often)? Uniformly recurrent (consecutive occurrences of a block are separated by bounded gaps)? Linearly recurrent? (Gap size between consecutive occurrences of length-n blocks is O(n)), etc.
The kinds of number-theoretic questions we are interested in include:

- given a set of integers $S$, can every element of $\mathbb{N}$ be written as the sum of $t$ elements of $S$?

(additive number theory)

- take an automatic sequence $(a_n)_{n \geq 0}$ and consider the real number $\sum_{n \geq 0} a_n k^{-n}$. Is it rational? Irrational? Algebraic? Transcendental? A Liouville number?

- given an automatic set, does it contain a prime number? Infinitely many primes?

- can one prove that $\pi$, $\log 2$, etc. do not have base-$k$ representations that are $k$-automatic?

- what is the asymptotic behavior of sums like

  $$\sum_{0 \leq n < N} S_2(n)$$

  $$\sum_{0 \leq n < N} R(n) \quad (Rudin-Shapiro)$$
Some notation we use throughout:

\((n)_k\) - canonical representation of \(n\) in base \(k, k \geq 2\) using digits 0, 1, ..., \(k-1\) and no leading zeroes. Note: \((0)_k = \varepsilon\), the empty word.

\([w]_k\) - number represented by \(w\) in base \(k, k \geq 2\) msd first. More precisely, if \(w = a_1a_2 \ldots a_i \ldots\) then

\[\[w\]_k = \sum_{1 \leq j \leq i} a_j k^{i-j}.\] (No restriction on \(a_i\).)

\(\gamma_k(n)\) - exponent of highest power of \(k\) dividing \(n\) (for \(n > 0, k \geq 2\)).

\(V_k(n) := k^{\gamma_k(n)}\)

\(W^R\) - reversal of word \(W\)

\((\text{drawer})^R = \text{reward}\)

\(\sum_k = \{0, 1, \ldots, k-1\}\) canonical base-\(k\) alphabet

\(\sum^k = \text{all finite words over alphabet } \sum\)

\(\sum^w = \text{all 1-sided infinite words over } \sum\)

\(\omega^w = \text{all 2-sided infinite words over } \sum\)

\(\omega = \text{the infinite word } \ldots \)
$|X|$ - length of string $X$

$|X|_a$ - number of occurrences of symbol $a$ in $X$

$X[i]$ - $i$'th letter of $X$

$X[i..j]$ - subword of $X$ starting at position $i$ and ending at position $j$
Course info:

- 3 problem sets + project
- Textbook *Automatic Sequences* available in bookstore
- Course home page:
  https://cs.uwaterloo.ca/~nshallit/Courses/860
  - Has lecture notes, problem sets, etc.
- Office hours 2:30-3:20 PM Wed
  Or by appointment
  Or when office door open