# Degrees of Streams

Jörg Endrullis Dimitri Hendriks
Jan Willem Klop

Vrije Universiteit Amsterdam

Streams Seminar
Nijmegen, 20th April 2010

# Complexity of streams

Complexity measures for infinite streams:

- Subword complexity
- Kolmogorov complexity

Comparing streams by transforming them into each other:

- Recursion theoretic degrees of unsolvability
  (transformation via Turing machines)

# Complexity of streams

Complexity measures for infinite streams:

- Subword complexity
- Kolmogorov complexity

Comparing streams by transforming them into each other:

- Recursion theoretic degrees of unsolvability (transformation via Turing machines)

We envisage:

- infinitary view on information content
- complexity invariant under exchange of finitely many elements
- capture the intrinsic, invariant infinite pattern of streams

We propose a comparison via finite state transducers (FSTs).

# Subword complexity

## Definition

Subword complexity is a measure on streams $\sigma$, that records as a function of $n$, how many of the finite words of length $n$ occur in $\sigma$.

Examples:

- Sturmian words: $n+1$
- morphic words: linear
- automatic sequences: quadratic

# Subword complexity

## Definition

Subword complexity is a measure on streams $\sigma$, that records as a function of $n$, how many of the finite words of length $n$ occur in $\sigma$.

Examples:

- Sturmian words: $n+1$
- morphic words: linear
- automatic sequences: quadratic

## Interesting, but...

Even non-computable streams can have linear subword complexity.

# Kolmogorov complexity

## Definition

The Kolmogorov complexity $\mathscr{K}(w)$ of a word $w$ is the length of the shortest program computing $w$.
(in a fixed universal programming system, e.g., Turing machines)

Examples:

- Thue–Morse: $\mathscr{K}(\mathrm{M}) \approx 6$ (Turing machine needs 6 states)

# Kolmogorov complexity

## Definition

The Kolmogorov complexity $\mathscr{K}(w)$ of a word $w$ is the length of the shortest program computing $w$.
(in a fixed universal programming system, e.g., Turing machines)

Examples:

- Thue–Morse: $\mathscr{K}(\mathrm{M}) \approx 6$ (Turing machine needs 6 states)

## Interesting, but...

The Kolmogorov complexity can be increased arbitrarily by:

- prefixing a finite word,
- changing the encoding
  ($0 \mapsto$ `I am a zero!`, $1 \mapsto$ `Here is a one!`)

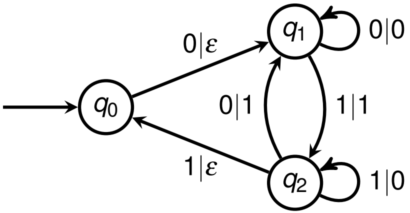
# Finite state transducers

## Definition

A finite state transducer (FST) is a deterministic finite automaton with:

- output words $w \in \Sigma^*$ along the edges,
- a transition function $\delta : Q \times \Sigma \to Q$,
- an output function $\lambda : Q \times \Sigma \to \Gamma^*$.

The following automaton computes the diff of a stream:




Thus it reduces Thue–Morse to Toeplitz:

$$01101001\ldots \to 1011101\ldots$$

# Partial order of stream degrees

## Definition (Equivalence of streams)

We write $M \triangleright N$ if there exists an FST that transforms $M$ into $N$.

$$\diamond := \triangleright \cap \triangleleft$$

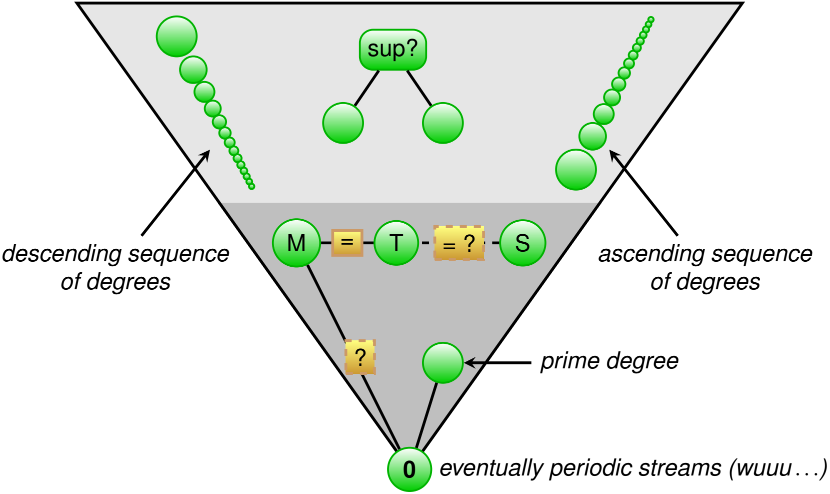
We use $\sigma^{\diamond} := \{\tau \mid \sigma \diamond \tau\}$ to denote the equivalence class of $\sigma$.

Note that:

- $\triangleright$ is reflexive and transitive ($\triangleright^* \subseteq \triangleright$)
- $\triangleright$ implies a partial order on the equivalence classes w.r.t. $\diamond$.

We are interested in the hierarchy of streams created by $\triangleright$.

# Hierarchy of streams




A stream $M$ is prime if there is no $N$ strictly in-between $M$ and 0.

# Hierarchy of streams: degrees are countable

We can enumerate all FSTs (and hence all reducts of a stream). Hence:

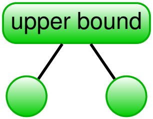
**Theorem**

Every degree is countable.

**Theorem**

Every degree has only a countable number of degrees below it.

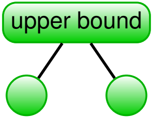
# Hierarchy of streams: upper bounds




**Lemma**

$$\mathrm{zip}_{n,m}(\sigma, \tau) \triangleright \sigma \qquad \mathrm{zip}_{n,m}(\sigma, \tau) \triangleright \tau$$

# Hierarchy of streams: upper bounds




## Lemma

$$\mathsf{zip}_{n,m}(\sigma,\tau) \rhd \sigma \qquad\qquad \mathsf{zip}_{n,m}(\sigma,\tau) \rhd \tau$$

## Theorem

A set $A$ of streams has an upper bound $\Longleftrightarrow$ $A$ is countable.

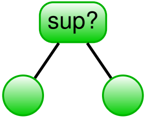
## Proof.

Let $A = \{\sigma_1, \sigma_2, \ldots\}$ be a set of streams. We define:

$$\tau_n = \mathsf{zip}(\sigma_n, \tau_{n+1})$$

Then $\tau_1$ is an upper bound, that is, $\tau_1 \rhd \sigma_n$ for all $n$. □

# Hierarchy of streams: least upper bounds

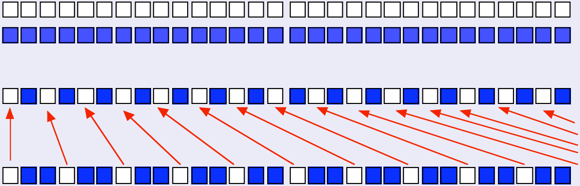



## Theorem

Not every pair of streams has a supremum.

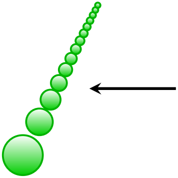
## Proof.

For suitable $\sigma$ and $\tau$: no common reduct of $\text{zip}_{1,1}(\sigma, \tau)$ and $\text{zip}_{1,2}(\sigma, \tau)$ is an upper bound for $\sigma$ and $\tau$.



□

# Hierarchy of streams: infinite ascending sequences




**Theorem**

There exist infinite ascending sequences.

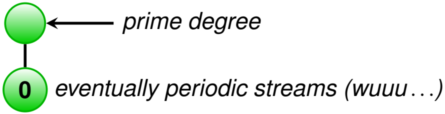
**Proof.**

Take any stream $\sigma$.

- The degree $\sigma^{\diamond}$ is countable.
- There exist uncountably many streams.

Hence there exists $\tau$ such that $\sigma \not\trianglerighteq \tau$.
Then $\text{zip}(\sigma, \tau) \triangleright \sigma$ but not $\sigma \triangleright \text{zip}(\sigma, \tau)$. $\square$

# Hierarchy of streams: primes




## Definition

A stream $M$ is prime if there exists no $N$ such that:

$$M \triangleright N \qquad\qquad 0 \ntriangleright N \qquad\qquad N \ntriangleright M,$$

that is $N$ is strictly in-between $M$ and 0.

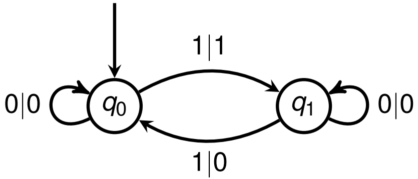
## Theorem

The following stream is prime:

$$\begin{aligned} P &= 101001000100001000001\ldots \\ &= 1\,0^1\,1\,0^2\,1\,0^3\,1\,0^4\,1\,0^5\,1\,0^6\,1\ldots \end{aligned}$$

# A prime stream: $P = 1101001000100001000001\ldots$

Heuristic evidence 1:

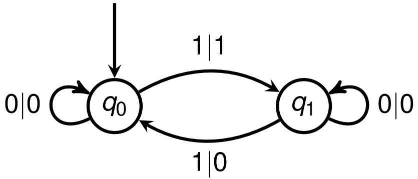



This FST deletes every second 1, that is, it reduces $P$ to:

$$\begin{aligned} P_1 &= 10000100000000100000000000001\ldots \\ &= 1\,0^4\,1\,0^8\,1\,0^{12}\,1\,0^{16}\,1\ldots \end{aligned}$$

# A prime stream: $P = 1101001000100001000001\ldots$

Heuristic evidence 1:




This FST deletes every second 1, that is, it reduces $P$ to:

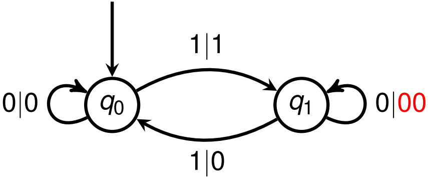
$$P_1 = 10000100000000100000000000001\ldots$$
$$= 1\,0^4\,1\,0^8\,1\,0^{12}\,1\,0^{16}\,1\ldots$$

We can transform $P_1$ back to $P$ by:

$$0000 \mapsto 0$$

# A prime stream: $P = 1101001000100001000001\ldots$

Heuristic evidence 2:

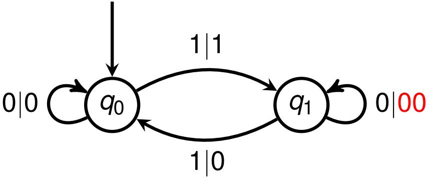



This FST deletes every second 1, that is, it reduces $P$ to:

$$P_2 = 1000001000000000001000000000000000001\ldots$$
$$= 1\,0^5\,1\,0^{11}\,1\,0^{17}\,1\ldots$$

# A prime stream: $P = 11010010001000001000001\ldots$

Heuristic evidence 2:




This FST deletes every second 1, that is, it reduces $P$ to:

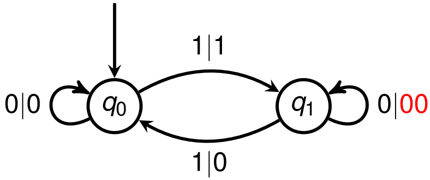
$$P_2 = 100000100000000000100000000000000000001\ldots$$
$$= 1\,0^5\,1\,0^{11}\,1\,0^{17}\,1\ldots$$

We can transform $P_2$ back to $P$ by compressing blocks of zeros:

$$0^n \mapsto 0^{\frac{n+1}{6}}$$

# A prime stream: $P = 1101001000100001000001\ldots$

Heuristic evidence 2:




This FST deletes every second 1, that is, it reduces $P$ to:

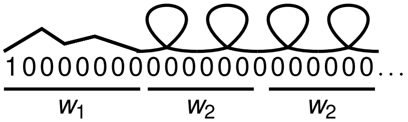
$$P_2 = 1000001000000000001000000000000000001\ldots$$
$$= 1\,0^5\,1\,0^{11}\,1\,0^{17}\,1\ldots$$

We can transform $P_2$ back to $P$ by compressing blocks of zeros:

$$0^n \mapsto 0^{\frac{n+1}{6}}$$

FSTs can perform arbitrary linear compressions.

# A prime stream: $P = 110100100010000010000001\ldots$




## Lemma

Let $Z$ be the least common multiple of all 0-loops in the FST.

For all $n > |Q|$ and states $s$ there exist $w_1, w_2 \in \Gamma^*$ s.t. for all $i \in \mathbb{N}$:

$$\delta(s, 10^{n+i \cdot Z}) = \delta(s, 10^n)$$

$$\lambda(s, 10^{n+i \cdot Z}) = w_1 w_2^i$$

## Proof.

Analogous to the pumping lemma for regular languages. □

# A prime stream: $P = 110100100001000001000001\ldots$

## Lemma

For every FST $A$ there exist $n \in \mathbb{N}$, $w, w_{j,1}, w_{j,2} \in \Gamma^*$ such that:

$$A(P) = w \cdot \prod_{i=0}^{\infty} \prod_{j=0}^{n} w_{j,1} \cdot w_{j,2}^{i}$$

## Proof.

By the pigeonhole principle we find blocks $10^{m_1}$ and $10^{m_2}$ in $P$ s.t.:

- $|Q| < m_1 < m_2$,
- $m_1 \equiv m_2 \mod Z$,
- the FST $A$ enters $10^{m_1}$ and $10^{m_2}$ with the same state $q$.

Define $n = m_2 - m_1$. Then we have:

- $A$ also leaves $10^{m_1}$ and $10^{m_2}$ with the same state $q'$, and
- $m_1 + 1 \equiv m_2 + 1 \mod Z, \ldots$

The $w_{j,1}, w_{j,2}$ are derived from the previous lemma. □

# Questions and open problems

- Is Sierpinsky interreducible with Morse?
- Is Morse prime?
- How many primes are out there?
- Are there interesting invariants for FST-transductions?