Departamento de Computación, Facultad de Ciencias Exactas y Naturales, UBA

Azar y Autómatas

Clase 6: Autómatas finitos y secuencias normales

Normality as incompressibility by finite automata

The definition of normality can be expressed as a notion of incompressibility by finite automata with output also known as transducers.

We focus on transducers that operate in real-time, that is, they process exactly one input alphabet symbol per transition. We consider non-deterministic transducers.

Non-deterministic real-time finite automata with output

A non-deterministic transducer is a tuple $\mathcal{T} = \langle Q, A, B, \delta, I, F \rangle$, where

- Q is a finite set of states,

Non-deterministic real-time finite automata with output

A non-deterministic transducer is a tuple $\mathcal{T} = \langle Q, A, B, \delta, I, F \rangle$, where

- Q is a finite set of states,
- A and B are the input and output alphabets, respectively.

Non-deterministic real-time finite automata with output

A non-deterministic transducer is a tuple $\mathcal{T} = \langle Q, A, B, \delta, I, F \rangle$, where

- Q is a finite set of states,
- A and B are the input and output alphabets, respectively.
- $\delta \subset Q \times A \times B^* \times Q$ is a transition relation.

Non-deterministic real-time finite automata with output

A non-deterministic transducer is a tuple $\mathcal{T} = \langle Q, A, B, \delta, I, F \rangle$, where

- Q is a finite set of states,
- A and B are the input and output alphabets, respectively.
- $\delta \subset Q \times A \times B^* \times Q$ is a transition relation. A transition is a tuple $\langle p, a, v, q \rangle$ which is written

$$p \xrightarrow{a|v} q$$

Non-deterministic real-time finite automata with output

A non-deterministic transducer is a tuple $\mathcal{T} = \langle Q, A, B, \delta, I, F \rangle$, where

- Q is a finite set of states,
- A and B are the input and output alphabets, respectively.
- $\delta \subset Q \times A \times B^* \times Q$ is a transition relation. A transition is a tuple $\langle p, a, v, q \rangle$ which is written

$$p \xrightarrow{a|v} q$$

- $I \subseteq Q$ and $F \subseteq Q$ are the sets of initial and final states, respectively.

Non-deterministic real-time finite automata with output

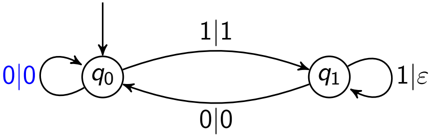
A non-deterministic transducer is a tuple $\mathcal{T} = \langle Q, A, B, \delta, I, F \rangle$, where

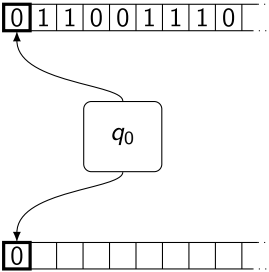
- Q is a finite set of states,
- A and B are the input and output alphabets, respectively.
- $\delta \subset Q \times A \times B^* \times Q$ is a transition relation. A transition is a tuple $\langle p, a, v, q \rangle$ which is written

$$p \xrightarrow{a|v} q$$

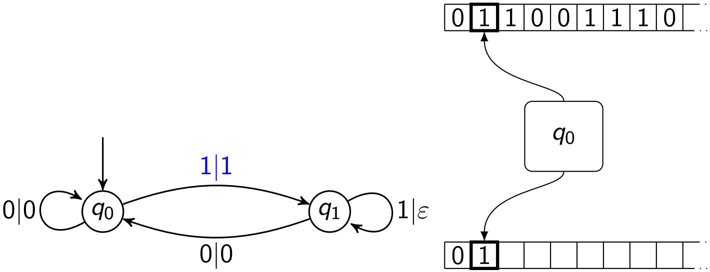
- $I \subseteq Q$ and $F \subseteq Q$ are the sets of initial and final states, respectively.

Finite automata with output



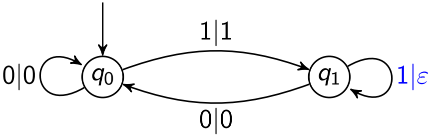


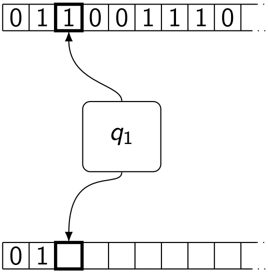
Finite automata with output



The transducer transforms rows of 1s into a single 1.

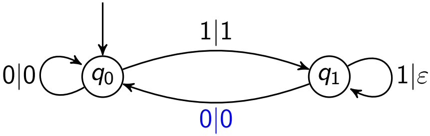
Finite automata with output

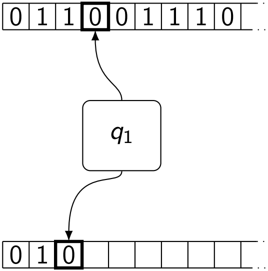




The transducer transforms rows of 1s into a single 1.

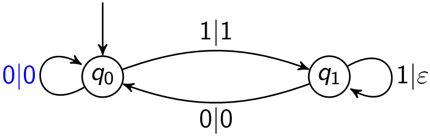
Finite automata with output

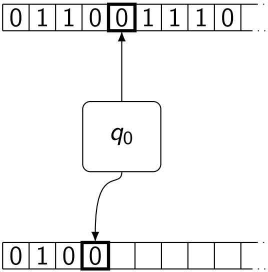




The transducer transforms rows of 1s into a single 1.

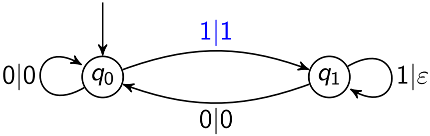
Finite automata with output

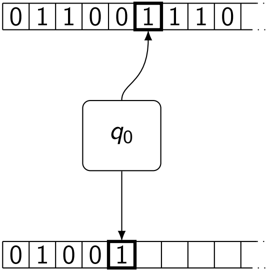




The transducer transforms rows of 1s into a single 1.

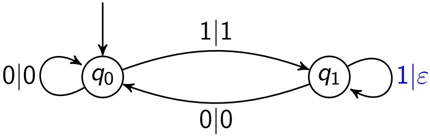
Finite automata with output

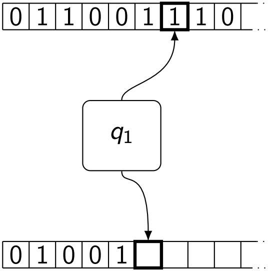




The transducer transforms rows of 1s into a single 1.

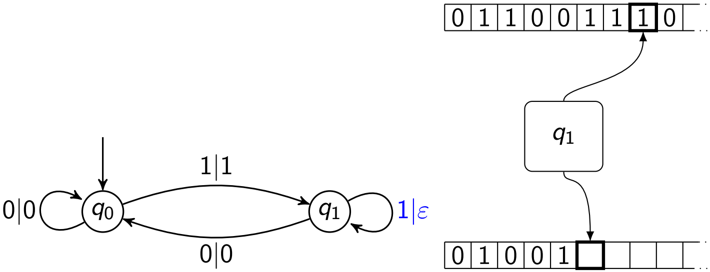
Finite automata with output





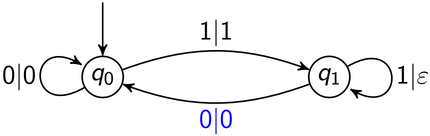
The transducer transforms rows of 1s into a single 1.

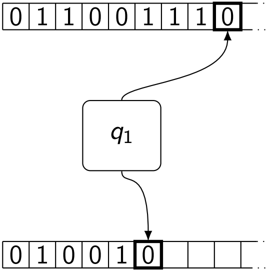
Finite automata with output



The transducer transforms rows of 1s into a single 1.

Finite automata with output





The transducer transforms rows of 1s into a single 1.

Finite automata with output

A run is a sequence of consecutive transitions,

$$q_0 \xrightarrow{a_1 | v_1} q_1 \xrightarrow{a_2 | v_2} q_2 \cdots q_{n-1} \xrightarrow{a_n | v_n} q_n \ldots$$

A finite run is written

$$q_0 \xrightarrow{a_1 \cdots a_n | v_1 \cdots v_n} q_n$$

Finite automata with output

A run is a sequence of consecutive transitions,

$$q_0 \xrightarrow{a_1|v_1} q_1 \xrightarrow{a_2|v_2} q_2 \cdots q_{n-1} \xrightarrow{a_n|v_n} q_n \ldots$$

A finite run is written

$$q_0 \xrightarrow{a_1\cdots a_n|v_1\cdots v_n} q_n$$

An infinite run is written

$$q_0 \xrightarrow{a_1 a_2 a_3 \cdots|v_1 v_2 v_3 \cdots} \infty$$

Finite automata with output

A run is a sequence of consecutive transitions,

$$q_0 \xrightarrow{a_1|v_1} q_1 \xrightarrow{a_2|v_2} q_2 \cdots q_{n-1} \xrightarrow{a_n|v_n} q_n \ldots$$

A finite run is written

$$q_0 \xrightarrow{a_1 \cdots a_n | v_1 \cdots v_n} q_n$$

An infinite run is written

$$q_0 \xrightarrow{a_1 a_2 a_3 \cdots | v_1 v_2 v_3 \cdots} \infty$$

An infinite run is accepting if q_0 is in I and for infinitely many ns q_n in F. This is the classical Büchi acceptance condition.

Finite automata with output

A run is a sequence of consecutive transitions,

$$q_0 \xrightarrow{a_1|v_1} q_1 \xrightarrow{a_2|v_2} q_2 \cdots q_{n-1} \xrightarrow{a_n|v_n} q_n \ldots$$

A finite run is written

$$q_0 \xrightarrow{a_1 \cdots a_n | v_1 \cdots v_n} q_n$$

An infinite run is written

$$q_0 \xrightarrow{a_1 a_2 a_3 \cdots | v_1 v_2 v_3 \cdots} \infty$$

An infinite run is accepting if q_0 is in I and for infinitely many ns q_n in F. This is the classical Büchi acceptance condition.

For two infinite words $x \in A^\omega$ and $y \in B^\omega$, we write $\mathcal{T}(x, y)$ whenever there is an accepting run $q_0 \xrightarrow{x|y} \infty$ in $\mathcal{T}$.

$\mathcal{T}$ is bounded-to-one if the function $y \mapsto |\{x : \mathcal{T}(x, y)\}|$ is bounded.

Compressibility by finite automata

An infinite word $x = a_1 a_2 a_3 \cdots$ is compressible if there is a bounded-to-one non-deterministic transducer with an accepting run $q_0 \xrightarrow{a_1 | v_1} q_1 \xrightarrow{a_2 | v_2} q_2 \xrightarrow{a_3 | v_3} q_3 \cdots$ satisfying

$$\liminf_{n \to \infty} \frac{|v_1 v_2 \cdots v_n|}{n} \frac{\log |B|}{\log |A|} < 1.$$

Example of a compressible sequence

The sequence $x = 01010101010101\ldots$ is compressible because there is a one-to-one automaton $\mathcal{T}$ that maps $0101 \to 0$ and for every v such that $|v| = 4$ and $v \neq 0101$ $v \to 1v$. Then,

Example of a compressible sequence

The sequence $x = 01010101010101\ldots$ is compressible because there is a one-to-one automaton $\mathcal{T}$ that maps $0101 \to 0$ and for every v such that $|v| = 4$ and $v \neq 0101$ $v \to 1v$. Then,

$$\liminf_{n\to\infty} \frac{|0^n|}{|(0101)^n|} = \liminf_{n\to\infty} \frac{n}{4n} = 1/4 < 1.$$

Example of a compressible sequence

The sequence $x = 0100100010000100000100000001\ldots$ is compressible because there is a one-to-one automaton $\mathcal{T}$ that maps $000 \to 0$ and for every v such that $|v| = 3$ and $v \neq 000$ $v \to 1v$. Then,

$$\liminf_{n\to\infty} \qquad \frac{|1010| + \left(\sum_{j=2}^{n} 4 + (j-2)/3 + 1\right)}{\sum_{j=1}^{n} j + 1}$$

Example of a compressible sequence

The sequence $x = 0100100010000100000100000001\ldots$ is compressible because there is a one-to-one automaton $\mathcal{T}$ that maps $000 \to 0$ and for every v such that $|v| = 3$ and $v \neq 000$ $v \to 1v$. Then,

$$\begin{aligned}
\liminf_{n\to\infty} \quad & \frac{|1010| + \left(\sum_{j=2}^{n} 4 + (j-2)/3 + 1\right)}{\sum_{j=1}^{n} j + 1} \\
< \quad & \liminf_{n\to\infty} \frac{4 + \sum_{j=2}^{n} 4 + \sum_{j=2}^{n} (j+1)/3}{\sum_{j=1}^{n} (j+1)}
\end{aligned}$$

Example of a compressible sequence

The sequence $x = 0100100010000100000100000001\ldots$ is compressible because there is a one-to-one automaton $\mathcal{T}$ that maps $000 \to 0$ and for every v such that $|v| = 3$ and $v \neq 000$ $v \to 1v$. Then,

$$\begin{aligned}
\liminf_{n\to\infty} \quad & \frac{|1010| + \left(\sum_{j=2}^{n} 4 + (j-2)/3 + 1\right)}{\sum_{j=1}^{n} j + 1} \\
< \quad & \liminf_{n\to\infty} \frac{4 + \sum_{j=2}^{n} 4 + \sum_{j=2}^{n} (j+1)/3}{\sum_{j=1}^{n} (j+1)} \\
< \quad & \liminf_{n\to\infty} \frac{4 + 4n}{n(n+1)/2} + 1/3
\end{aligned}$$

Example of a compressible sequence

The sequence $x = 0100100010000100000100000001\ldots$ is compressible because there is a one-to-one automaton $\mathcal{T}$ that maps $000 \to 0$ and for every v such that $|v| = 3$ and $v \neq 000$ $v \to 1v$. Then,

$$
\begin{aligned}
\liminf_{n\to\infty} \quad & \frac{|1010| + \left(\sum_{j=2}^{n} 4 + (j-2)/3 + 1\right)}{\sum_{j=1}^{n} j + 1} \\
< \quad & \liminf_{n\to\infty} \frac{4 + \sum_{j=2}^{n} 4 + \sum_{j=2}^{n} (j+1)/3}{\sum_{j=1}^{n} (j+1)} \\
< \quad & \liminf_{n\to\infty} \frac{4+4n}{n(n+1)/2} + 1/3 \\
= \quad & 1/3 \\
< \quad & 1.
\end{aligned}
$$

Normality as incompressibility by finite automata

Theorem 1

An infinite word is normal if and only if it is not compressible by any bounded-to-one non-deterministic transducer.

Normal implies not compressible

Lemma 2

Let ℓ be a positive integer and let $u_1, u_2, u_3, \ldots$ be words of length ℓ over the alphabet A such that $x = u_1 u_2 u_3 \cdots$ is simply normal to length ℓ. Consider a bounded-to-one transducer and let an accepting run with input x

$$q_0 \xrightarrow{u_1|v_1} q_1 \xrightarrow{u_2|v_2} q_2 \xrightarrow{u_3|v_3} q_3 \cdots$$

Assume there is a real $\varepsilon > 0$ and a set $U \subseteq A^\ell$ of at least $(1-\varepsilon)|A|^\ell$ words such that $u_i \in U$ implies $|v_i| \geq \ell(1-\varepsilon)$. Then,

$$\liminf_{n\to\infty} \frac{|v_1 v_2 \cdots v_n|}{n\ell} \geq (1-\varepsilon)^3.$$

Proof of Lemma 2

By hypothesis $x = u_1 u_2 \ldots$, where each $|u_i| = \ell$, is simply normal to length ℓ.

Proof of Lemma 2

By hypothesis $x = u_1 u_2 \ldots$, where each $|u_i| = \ell$, is simply normal to length ℓ.Let n_0 be such that for every $u \in A^\ell$ and for every $n \geq n_0$,

$$|\{i : 1 \leq i \leq n, u_i = u\}| \geq \frac{n}{|A|^\ell}(1 - \varepsilon).$$

Proof of Lemma 2

By hypothesis $x = u_1 u_2 \ldots$, where each $|u_i| = \ell$, is simply normal to length ℓ.Let n_0 be such that for every $u \in A^\ell$ and for every $n \geq n_0$,

$$|\{i : 1 \leq i \leq n, u_i = u\}| \geq \frac{n}{|A|^\ell}(1 - \varepsilon).$$

Then, for every $n \geq n_0$,

$$|v_1 v_2 \cdots v_n| \quad = \quad \sum_{i=1}^{n} |v_i|$$

Proof of Lemma 2

By hypothesis $x = u_1 u_2 \ldots$, where each $|u_i| = \ell$, is simply normal to length ℓ.Let n_0 be such that for every $u \in A^\ell$ and for every $n \geq n_0$,

$$|\{i : 1 \leq i \leq n, u_i = u\}| \geq \frac{n}{|A|^\ell}(1 - \varepsilon).$$

Then, for every $n \geq n_0$,

$$\begin{aligned} |v_1 v_2 \cdots v_n| &= \sum_{i=1}^{n} |v_i| \\ &\geq \sum_{1 \leq i \leq n, u_i \in U} |v_i| \end{aligned}$$

Proof of Lemma 2

By hypothesis $x = u_1 u_2 \ldots$, where each $|u_i| = \ell$, is simply normal to length ℓ.Let n_0 be such that for every $u \in A^\ell$ and for every $n \geq n_0$,

$$|\{i : 1 \leq i \leq n, u_i = u\}| \geq \frac{n}{|A|^\ell}(1 - \varepsilon).$$

Then, for every $n \geq n_0$,

$$\begin{aligned}
|v_1 v_2 \cdots v_n| &= \sum_{i=1}^{n} |v_i| \\
&\geq \sum_{1 \leq i \leq n, u_i \in U} |v_i| \\
&\geq \sum_{1 \leq i \leq n, u_i \in U} \ell(1 - \varepsilon)
\end{aligned}$$

Proof of Lemma 2

By hypothesis $x = u_1 u_2 \ldots$, where each $|u_i| = \ell$, is simply normal to length ℓ.Let n_0 be such that for every $u \in A^\ell$ and for every $n \geq n_0$,

$$|\{i : 1 \leq i \leq n, u_i = u\}| \geq \frac{n}{|A|^\ell}(1 - \varepsilon).$$

Then, for every $n \geq n_0$,

$$\begin{aligned}
|v_1 v_2 \cdots v_n| &= \sum_{i=1}^{n} |v_i| \\
&\geq \sum_{1 \leq i \leq n, u_i \in U} |v_i| \\
&\geq \sum_{1 \leq i \leq n, u_i \in U} \ell(1 - \varepsilon) \\
&\geq n|A|^{-\ell}(1 - \varepsilon) \sum_{u \in U} \ell(1 - \varepsilon)
\end{aligned}$$

Proof of Lemma 2

By hypothesis $x = u_1 u_2 \ldots$, where each $|u_i| = \ell$, is simply normal to length ℓ.Let n_0 be such that for every $u \in A^\ell$ and for every $n \geq n_0$,

$$|\{i : 1 \leq i \leq n, u_i = u\}| \geq \frac{n}{|A|^\ell}(1 - \varepsilon).$$

Then, for every $n \geq n_0$,

$$\begin{aligned}
|v_1 v_2 \cdots v_n| &= \sum_{i=1}^{n} |v_i| \\
&\geq \sum_{1 \leq i \leq n, u_i \in U} |v_i| \\
&\geq \sum_{1 \leq i \leq n, u_i \in U} \ell(1 - \varepsilon) \\
&\geq n|A|^{-\ell}(1 - \varepsilon) \sum_{u \in U} \ell(1 - \varepsilon) \\
&\geq n|A|^{-\ell}(1 - \varepsilon)(1 - \varepsilon)|A|^\ell \ell(1 - \varepsilon)
\end{aligned}$$

Proof of Lemma 2

By hypothesis $x = u_1 u_2 \ldots$, where each $|u_i| = \ell$, is simply normal to length ℓ.Let n_0 be such that for every $u \in A^\ell$ and for every $n \geq n_0$,

$$|\{i : 1 \leq i \leq n, u_i = u\}| \geq \frac{n}{|A|^\ell}(1 - \varepsilon).$$

Then, for every $n \geq n_0$,

$$\begin{aligned}
|v_1 v_2 \cdots v_n| &= \sum_{i=1}^{n} |v_i| \\
&\geq \sum_{1 \leq i \leq n, u_i \in U} |v_i| \\
&\geq \sum_{1 \leq i \leq n, u_i \in U} \ell(1 - \varepsilon) \\
&\geq n|A|^{-\ell}(1 - \varepsilon) \sum_{u \in U} \ell(1 - \varepsilon) \\
&\geq n|A|^{-\ell}(1 - \varepsilon)(1 - \varepsilon)|A|^{\ell} \ell(1 - \varepsilon) \\
&\geq (1 - \varepsilon)^3 n\ell. \qquad \square
\end{aligned}$$

Normal implies not compressible

Lemma 3

If x is normal then x can not be compressed by any bounded-to-one real-time transducer.

Proof of Lemma 3

Fix a normal infinite word $x = a_1 a_2 a_3 \cdots$, a real $\varepsilon > 0$, a bounded-to-one non-deterministic $\mathcal{T} = \langle Q, A, B, \delta, q_0, F \rangle$ and an accepting run

$$q_0 \xrightarrow{a_1|v_1} q_1 \xrightarrow{a_2|v_2} q_2 \xrightarrow{a_3|v_3} q_3 \cdots$$

where $a \in A$ and $v \in B^*$.

Proof of Lemma 3

Fix a normal infinite word $x = a_1 a_2 a_3 \cdots$, a real $\varepsilon > 0$, a bounded-to-one non-deterministic $\mathcal{T} = \langle Q, A, B, \delta, q_0, F \rangle$ and an accepting run

$$q_0 \xrightarrow{a_1 | v_1} q_1 \xrightarrow{a_2 | v_2} q_2 \xrightarrow{a_3 | v_3} q_3 \cdots$$

where $a \in A$ and $v \in B^*$.

It suffices to show that there is ℓ and U such that Lemma 2 applies to this arbitrary choice of ε, $\mathcal{T}$ and accepting run.

Proof of Lemma 3

For each length ℓ, pair of states p, q in the run and each word v define

$$U'(\ell, p, q, v) = \{u \in A^\ell : p \xrightarrow{u|v} q\}.$$

Proof of Lemma 3

For each length ℓ, pair of states p, q in the run and each word v define

$$U'(\ell, p, q, v) = \{u \in A^\ell : p \xrightarrow{u|v} q\}.$$

Thus, for each $u \in U'(\ell, p, q, v)$ we can write

$$q_0 \xrightarrow{x|y} \infty = q_0 \xrightarrow{u_0|v_0} p \xrightarrow{u|v} q \xrightarrow{x_0|y_0} \infty$$

Proof of Lemma 3

For each length ℓ, pair of states p, q in the run and each word v define

$$U'(\ell, p, q, v) = \{u \in A^\ell : p \xrightarrow{u|v} q\}.$$

Thus, for each $u \in U'(\ell, p, q, v)$ we can write

$$q_0 \xrightarrow{x|y} \infty = q_0 \xrightarrow{u_0|v_0} p \xrightarrow{u|v} q \xrightarrow{x_0|y_0} \infty$$

Let t be such that T is t-to-one. Observe that $|U'(\ell, p, q, v)| \leq t$.

Proof of Lemma 3

For each length ℓ, pair of states p, q in the run and each word v define

$$U'(\ell, p, q, v) = \{u \in A^\ell : p \xrightarrow{u|v} q\}.$$

Thus, for each $u \in U'(\ell, p, q, v)$ we can write

$$q_0 \xrightarrow{x|y} \infty = q_0 \xrightarrow{u_0|v_0} p \xrightarrow{u|v} q \xrightarrow{x_0|y_0} \infty$$

Let t be such that T is t-to-one. Observe that $|U'(\ell, p, q, v)| \leq t$.

For each $u \in A^*$ let h_u be its minimum output length

$$h_u = \min\{|v| : \exists i, j, 0 \leq i \leq j, q_i \xrightarrow{u|v} q_j\}$$

Proof of Lemma 3

For each length ℓ, pair of states p, q in the run and each word v define

$$U'(\ell, p, q, v) = \{u \in A^\ell : p \xrightarrow{u|v} q\}.$$

Thus, for each $u \in U'(\ell, p, q, v)$ we can write

$$q_0 \xrightarrow{x|y} \infty = q_0 \xrightarrow{u_0|v_0} p \xrightarrow{u|v} q \xrightarrow{x_0|y_0} \infty$$

Let t be such that T is t-to-one. Observe that $|U'(\ell, p, q, v)| \leq t$.

For each $u \in A^*$ let h_u be its minimum output length

$$h_u = \min\{|v| : \exists i, j, 0 \leq i \leq j, q_i \xrightarrow{u|v} q_j\}$$

Let U_ℓ be the set of words of length ℓ that yield large output

$$U_\ell = \{u \in A^\ell : h_u \geq (1 - \varepsilon)\ell\}$$

Proof of Lemma 3

For each length ℓ, pair of states p, q in the run and each word v define

$$U'(\ell, p, q, v) = \{u \in A^\ell : p \xrightarrow{u|v} q\}.$$

Thus, for each $u \in U'(\ell, p, q, v)$ we can write

$$q_0 \xrightarrow{x|y} \infty = q_0 \xrightarrow{u_0|v_0} p \xrightarrow{u|v} q \xrightarrow{x_0|y_0} \infty$$

Let t be such that T is t-to-one. Observe that $|U'(\ell, p, q, v)| \leq t$.

For each $u \in A^*$ let h_u be its minimum output length

$$h_u = \min\{|v| : \exists i, j, 0 \leq i \leq j, q_i \xrightarrow{u|v} q_j\}$$

Let U_ℓ be the set of words of length ℓ that yield large output

$$U_\ell = \{u \in A^\ell : h_u \geq (1 - \varepsilon)\ell\}$$

Then,

$$|U_\ell| \geq |A|^\ell - |Q|^2 t |B|^{(1-\varepsilon)\ell+1}.$$

Proof of Lemma 3

For each length ℓ, pair of states p, q in the run and each word v define

$$U'(\ell, p, q, v) = \{u \in A^\ell : p \xrightarrow{u|v} q\}.$$

Thus, for each $u \in U'(\ell, p, q, v)$ we can write

$$q_0 \xrightarrow{x|y} \infty = q_0 \xrightarrow{u_0|v_0} p \xrightarrow{u|v} q \xrightarrow{x_0|y_0} \infty$$

Let t be such that T is t-to-one. Observe that $|U'(\ell, p, q, v)| \leq t$.

For each $u \in A^*$ let h_u be its minimum output length

$$h_u = \min\{|v| : \exists i, j, 0 \leq i \leq j, q_i \xrightarrow{u|v} q_j\}$$

Let U_ℓ be the set of words of length ℓ that yield large output

$$U_\ell = \{u \in A^\ell : h_u \geq (1 - \varepsilon)\ell\}$$

Then,

$$|U_\ell| \geq |A|^\ell - |Q|^2 t |B|^{(1-\varepsilon)\ell+1}.$$

Then, there is ℓ large enough such that $|U_\ell| > |A|^\ell(1 - \varepsilon)$ and apply Lemma 2 with $U = U_\ell$ to the considered run. □

Non-compressible implies normal

We show non-normal implies compressible.

Lemma 4

Every non-normal infinite word is compressible by some deterministic one-to-one transducer.

Non-compressible implies normal

We show non-normal implies compressible.

Lemma 4

Every non-normal infinite word is compressible by some deterministic one-to-one transducer.

This is stronger than we need (non-deterministic, bounded-to-one)

Proof of Lemma 4

Assume $x \in A^\omega$ is not normal.

Proof of Lemma 4

Assume $x \in A^\omega$ is not normal. Then, there is u_0 of length k such that

$$\lim_{n\to\infty} \frac{\|x[1\dots n]\|_{u_0}}{n/k} \neq \frac{1}{|A|^k}$$

Proof of Lemma 4

Assume $x \in A^\omega$ is not normal. Then, there is u_0 of length k such that

$$\lim_{n\to\infty} \frac{\|x[1\ldots n]\|_{u_0}}{n/k} \neq \frac{1}{|A|^k}$$

There exists then an increasing sequence $(n_i)_{i\geq 0}$ of integers such that for each word u of length k

$$f_u = \lim_{i\to\infty} \frac{\|x[1\ldots n_i]\|_u}{n_i/k}$$

and

Proof of Lemma 4

Assume $x \in A^\omega$ is not normal. Then, there is u_0 of length k such that

$$\lim_{n\to\infty} \frac{\|x[1\ldots n]\|_{u_0}}{n/k} \neq \frac{1}{|A|^k}$$

There exists then an increasing sequence $(n_i)_{i\geq 0}$ of integers such that for each word u of length k

$$f_u = \lim_{i\to\infty} \frac{\|x[1\ldots n_i]\|_u}{n_i/k}$$

and

$$f_{u_0} \neq \frac{1}{|A|^k}.$$

Note that $\sum_{u\in A^k} f_u = 1$.

Proof of Lemma 4

Let m be an integer to be fixed later.

Proof of Lemma 4

Let m be an integer to be fixed later. For each word $w \in A^{km}$ consider

$$w = u_1 \cdots u_m \text{ with } |u_i| = k \text{ for each } 1 \leq i \leq m$$

Proof of Lemma 4

Let m be an integer to be fixed later. For each word $w \in A^{km}$ consider

$$w = u_1 \cdots u_m \text{ with } |u_i| = k \text{ for each } 1 \leq i \leq m$$

and let

$$f_w = \prod_{i=1}^{m} f_{u_i}$$

Notice that $\sum_{w \in A^{km}} f_w = 1$.

Proof of Lemma 4

Let m be an integer to be fixed later. For each word $w \in A^{km}$ consider

$$w = u_1 \cdots u_m \text{ with } |u_i| = k \text{ for each } 1 \le i \le m$$

and let

$$f_w = \prod_{i=1}^{m} f_{u_i}$$

Notice that $\sum_{w \in A^{km}} f_w = 1$.
Now put A^{km} in a one to one correspondence with a prefix free set in B^*:
for each $w \in A^{km}$ let $v_w \in B^*$ such that

$$|v_w| \le \left\lceil \frac{-\log f_w}{\log |B|} \right\rceil .$$

Proof of Lemma 4

We construct a deterministic transducer $\mathcal{T}_m = \langle Q_m, A, B, \delta_m, I, F_m \rangle$,

$$\begin{aligned} Q_m &= A^{<km} \\ I &= \{\lambda\} \\ F_m &= Q_m \\ \delta_m &= \{w \xrightarrow{a|\lambda} wa : |wa| < km\} \cup \{w \xrightarrow{a|v_{wa}} \lambda : |wa| = km\}. \end{aligned}$$

The transducer $\mathcal{T}_m$ always comes back to its initial state λ after reading km symbols.

Let $\mathcal{T}_m(z)$ be the output of $\mathcal{T}_m$ on some finite input word z.

Let $\mathcal{T}_m(z)$ be the output of $\mathcal{T}_m$ on some finite input word z.
Factorize $z \in A^*$ as $z = w_1 \cdots w_n w'$ where $|w_i| = km$ for each $1 \leq i \leq n$ and $|w'| < km$. Note that $n = \lfloor |z|/(km) \rfloor$.

$$|\mathcal{T}_m(z)| = \sum_{i=1}^{n} |v_{w_i}|$$

Let $\mathcal{T}_m(z)$ be the output of $\mathcal{T}_m$ on some finite input word z.
Factorize $z \in A^*$ as $z = w_1 \cdots w_n w'$ where $|w_i| = km$ for each $1 \le i \le n$ and $|w'| < km$. Note that $n = \lfloor |z|/(km) \rfloor$.

$$
\begin{aligned}
|\mathcal{T}_m(z)| &= \sum_{i=1}^{n} |v_{w_i}| \\
&\le \sum_{i=1}^{n} \lceil - \log f_{w_i} / \log |B| \rceil
\end{aligned}
$$

Let $\mathcal{T}_m(z)$ be the output of $\mathcal{T}_m$ on some finite input word z.
Factorize $z \in A^*$ as $z = w_1 \cdots w_n w'$ where $|w_i| = km$ for each $1 \le i \le n$ and $|w'| < km$. Note that $n = \lfloor |z|/(km) \rfloor$.

$$
\begin{aligned}
|\mathcal{T}_m(z)| &= \sum_{i=1}^{n} |v_{w_i}| \\
&\le \sum_{i=1}^{n} \lceil -\log f_{w_i} / \log |B| \rceil \\
&\le \frac{|z|}{km} + \sum_{i=1}^{n} -\log f_{w_i} / \log |B|
\end{aligned}
$$

Let $\mathcal{T}_m(z)$ be the output of $\mathcal{T}_m$ on some finite input word z.
Factorize $z \in A^*$ as $z = w_1 \cdots w_n w'$ where $|w_i| = km$ for each $1 \leq i \leq n$ and $|w'| < km$. Note that $n = \lfloor |z|/(km) \rfloor$.

$$\begin{aligned}
|\mathcal{T}_m(z)| &= \sum_{i=1}^{n} |v_{w_i}| \\
&\leq \sum_{i=1}^{n} \lceil -\log f_{w_i} / \log |B| \rceil \\
&\leq \frac{|z|}{km} + \sum_{i=1}^{n} -\log f_{w_i} / \log |B| \\
&\leq \frac{|z|}{km} + \sum_{w \in A^{km}} \|z\|_w (-\log f_w)/\log |B|
\end{aligned}$$

Let $\mathcal{T}_m(z)$ be the output of $\mathcal{T}_m$ on some finite input word z.
Factorize $z \in A^*$ as $z = w_1 \cdots w_n w'$ where $|w_i| = km$ for each $1 \leq i \leq n$ and $|w'| < km$. Note that $n = \lfloor |z|/(km) \rfloor$.

$$
\begin{aligned}
|\mathcal{T}_m(z)| &= \sum_{i=1}^{n} |v_{w_i}| \\
&\leq \sum_{i=1}^{n} \lceil -\log f_{w_i} / \log |B| \rceil \\
&\leq \frac{|z|}{km} + \sum_{i=1}^{n} -\log f_{w_i} / \log |B| \\
&\leq \frac{|z|}{km} + \sum_{w \in A^{km}} \|z\|_w (-\log f_w)/\log |B| \\
&\leq \frac{|z|}{km} + \sum_{u \in A^{k}} \|z\|_u (-\log f_u)/\log |B|.
\end{aligned}
$$

Proof of Lemma 4

Applying this computation to the prefix $z = x[1..n]$ of x gives

Proof of Lemma 4

Applying this computation to the prefix $z = x[1..n]$ of x gives

$$\liminf_{n\to\infty} \frac{|\mathcal{T}_m(x[1..n])|}{n} \frac{\log|B|}{\log|A|} \leq \quad \lim_{i\to\infty} \frac{|\mathcal{T}_m(x[1..n_i])|}{n_i} \frac{\log|B|}{\log|A|}$$

Proof of Lemma 4

Applying this computation to the prefix $z = x[1..n]$ of x gives

$$\begin{aligned}
\liminf_{n\to\infty} \frac{|\mathcal{T}_m(x[1..n])|}{n}\frac{\log|B|}{\log|A|} &\leq \lim_{i\to\infty} \frac{|\mathcal{T}_m(x[1..n_i])|}{n_i}\frac{\log|B|}{\log|A|} \\
&\leq \lim_{i\to\infty} \frac{n_i}{n_i km}\frac{\log|B|}{\log|A|} + \sum_{u\in A^k} \frac{\|z\|_u}{k\ \ n_i/k}\frac{-\log f_u}{\log|B|}\frac{\log|B|}{\log|A|}
\end{aligned}$$

Proof of Lemma 4

Applying this computation to the prefix $z = x[1..n]$ of x gives

$$\begin{aligned}
\liminf_{n\to\infty} \frac{|\mathcal{T}_m(x[1..n])|}{n}\frac{\log|B|}{\log|A|} &\leq \lim_{i\to\infty} \frac{|\mathcal{T}_m(x[1..n_i])|}{n_i}\frac{\log|B|}{\log|A|} \\
&\leq \lim_{i\to\infty} \frac{n_i}{n_i km}\frac{\log|B|}{\log|A|} + \sum_{u\in A^k} \frac{\|z\|_u}{k\ \ n_i/k}\frac{-\log f_u}{\log|B|}\frac{\log|B|}{\log|A|} \\
&\leq \frac{1}{km}\frac{\log|B|}{\log|A|} + \frac{1}{k\log|A|}\sum_{u\in A^k} f_u(-\log f_u)
\end{aligned}$$

Proof of Lemma 4

Applying this computation to the prefix $z = x[1..n]$ of x gives

$$\begin{aligned}
\liminf_{n\to\infty} \frac{|\mathcal{T}_m(x[1..n])|}{n}\frac{\log|B|}{\log|A|} &\le \lim_{i\to\infty} \frac{|\mathcal{T}_m(x[1..n_i])|}{n_i}\frac{\log|B|}{\log|A|} \\
&\le \lim_{i\to\infty} \frac{n_i}{n_i km}\frac{\log|B|}{\log|A|} + \sum_{u\in A^k} \frac{\|z\|_u}{k\ \ n_i/k}\frac{-\log f_u}{\log|B|}\frac{\log|B|}{\log|A|} \\
&\le \frac{1}{km}\frac{\log|B|}{\log|A|} + \frac{1}{k\log|A|}\sum_{u\in A^k} f_u(-\log f_u) \\
&< \frac{1}{km}\frac{\log|B|}{\log|A|} + 1.
\end{aligned}$$

Proof of Lemma 4

Applying this computation to the prefix $z = x[1..n]$ of x gives

$$\begin{aligned}
\liminf_{n\to\infty} \frac{|\mathcal{T}_m(x[1..n])|}{n}\frac{\log|B|}{\log|A|} &\le \lim_{i\to\infty} \frac{|\mathcal{T}_m(x[1..n_i])|}{n_i}\frac{\log|B|}{\log|A|} \\
&\le \lim_{i\to\infty} \frac{n_i}{n_i km}\frac{\log|B|}{\log|A|} + \sum_{u\in A^k} \frac{\|z\|_u}{k\ \ n_i/k}\frac{-\log f_u}{\log|B|}\frac{\log|B|}{\log|A|} \\
&\le \frac{1}{km}\frac{\log|B|}{\log|A|} + \frac{1}{k\log|A|}\sum_{u\in A^k} f_u(-\log f_u) \\
&< \frac{1}{km}\frac{\log|B|}{\log|A|} + 1.
\end{aligned}$$

Since at least one number f_u is not equal to $1/|A|^k$,

$$\sum_{u\in A^k} f_u(-\log f_u) < k\log|A|.$$

Proof of Lemma 4

Applying this computation to the prefix $z = x[1..n]$ of x gives

$$\begin{aligned}
\liminf_{n\to\infty} \frac{|\mathcal{T}_m(x[1..n])|}{n}\frac{\log|B|}{\log|A|} &\le \lim_{i\to\infty} \frac{|\mathcal{T}_m(x[1..n_i])|}{n_i}\frac{\log|B|}{\log|A|} \\
&\le \lim_{i\to\infty} \frac{n_i}{n_i km}\frac{\log|B|}{\log|A|} + \sum_{u\in A^k} \frac{\|z\|_u}{k\ \ n_i/k}\frac{-\log f_u}{\log|B|}\frac{\log|B|}{\log|A|} \\
&\le \frac{1}{km}\frac{\log|B|}{\log|A|} + \frac{1}{k\log|A|}\sum_{u\in A^k} f_u(-\log f_u) \\
&< \frac{1}{km}\frac{\log|B|}{\log|A|} + 1.
\end{aligned}$$

Since at least one number f_u is not equal to $1/|A|^k$,

$$\sum_{u\in A^k} f_u(-\log f_u) < k\log|A|.$$

Then, for m chosen large enough, we obtain that $\mathcal{T}_m$ compresses x. □

Normality and finite automata

Theorem 5

normality   *iff*   *no finite-state martingale success*
(Schnorr and Stimm 1971)

Normality and finite automata

Theorem 5

normality	*iff*	*no finite-state martingale success* (Schnorr and Stimm 1971)
no finite-state martingale success	*iff*	*incompressibility* (Dai, Lathrop, Lutz and Mayordomo 2004) (Bourke, Hitchcock and Vinodchandran 2005)

Normality and finite automata

Theorem 5

normality	*iff*	*no finite-state martingale success* (Schnorr and Stimm 1971)
no finite-state martingale success	*iff*	*incompressibility* (Dai, Lathrop, Lutz and Mayordomo 2004) (Bourke, Hitchcock and Vinodchandran 2005)
normality	*iff*	*incompressibility (direct proof)* (Becher and Heiber 2013)

Normality and finite automata

Theorem 5

normality	*iff*	*no finite-state martingale success* (Schnorr and Stimm 1971)
no finite-state martingale success	*iff*	*incompressibility* (Dai, Lathrop, Lutz and Mayordomo 2004) (Bourke, Hitchcock and Vinodchandran 2005)
normality	*iff*	*incompressibility (direct proof)* (Becher and Heiber 2013)
	iff	*incompressibility by non-deterministic*
	iff	*incompressibility by one counter* (Becher, Carton and Heiber 2015)

Normality and finite automata

Theorem 5

normality	*iff*	*no finite-state martingale success* (Schnorr and Stimm 1971)
no finite-state martingale success	*iff*	*incompressibility* (Dai, Lathrop, Lutz and Mayordomo 2004) (Bourke, Hitchcock and Vinodchandran 2005)
normality	*iff*	*incompressibility (direct proof)* (Becher and Heiber 2013)
	iff	*incompressibility by non-deterministic*
	iff	*incompressibility by one counter* (Becher, Carton and Heiber 2015)
	iff	*incompressibility two-way transducers* (Carton and Heiber 2015)