Entropy of tree automata, joint spectral radii and zero-error coding

Cătălin Dima

LACL, Univ. Paris-Est

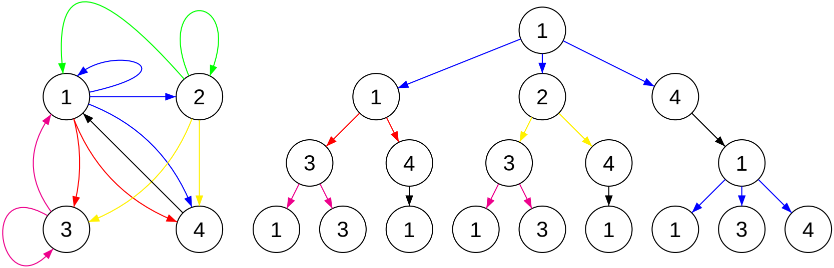
Automata Theory and Symbolic Dynamics Workshop, 3-7 June, 2013

Outline of the talk

- Entropy of trees & tree automata.
- Entropy of tree automata = joint spectral radius.
- The zero-error coding problem with states.
- Approximating zero-error capacities of codes.
- Open problems, etc.

Tree automata

- Tree automaton : $\mathcal{A} = (Q, \delta, q_0)$:
  - $\delta \subseteq Q \times 2^Q$.
  - Labels = states.
  - No accepting condition.
- (Infinite) Q-trees : $t : \mathbb{N}^* \to Q$ with prefix-closed domain.



- $L(\mathcal{A}, Acc) \neq \emptyset$ iff $\mathcal{A}$ accepts a regular tree.

Entropy of tree automata

- Entropy of a Q-tree :

$$\mathcal{H}(t) = \limsup_n \frac{1}{n} \log card(t\big|_n)$$

  - $t\big|_n$ = the set of nodes on level n.
  - Trees regarded as sets of runs sharing common prefixes.
- Entropy of a tree automaton :

$$\mathcal{H}(\mathcal{A}) = \sup \left\{ \mathcal{H}(t) \mid t \in \mathcal{L}(\mathcal{A}) \right\}$$

  - The size of the largest set of runs generated by $\mathcal{A}$.

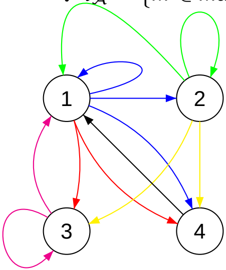
The set of adjacency matrices of a tree automaton

- Consider $Q = \{1, \ldots, n\}$.
- For each $i \leq n$ and $D \in \delta(i)$, put $m_{i,D} \in \{0,1\}^n$ the vector with :

$$m_{i,D}[j] = \begin{cases} 1 & \text{if } j \in D \\ 0 & \text{otherwise} \end{cases}$$

- The set of adjacency matrices over $\big(\delta(i)\big)_{1 \leq i \leq n}$ is

$$\mathcal{M}_{\mathcal{A}} = \big\{ M \in Mat(n \times n) \mid \exists D \text{ with } M[i, \bullet] = m_{i,D} \big\}$$



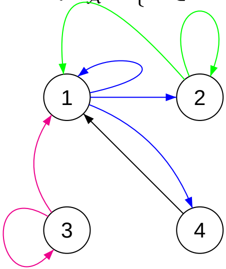
The set of adjacency matrices of a tree automaton

- Consider $Q = \{1, \dots, n\}$.
- For each $i \leq n$ and $D \in \delta(i)$, put $m_{i,D} \in \{0,1\}^n$ the vector with :

$$m_{i,D}[j] = \begin{cases} 1 & \text{if } j \in D \\ 0 & \text{otherwise} \end{cases}$$

- The set of adjacency matrices over $\left(\delta(i)\right)_{1 \leq i \leq n}$ is

$$\mathcal{M}_{\mathcal{A}} = \left\{ M \in Mat(n \times n) \mid \exists D \text{ with } M[i, \bullet] = m_{i,D} \right\}$$



$$A_1 = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

A property of the joint spectral radius

- The joint spectral radius of a set of matrices $\mathcal{N}$:

$$\rho(\mathcal{N}) = \sup_n \left\{ \|M_1 \cdot \ldots \cdot M_n\|^{1/n} \mid M_i \in \mathcal{N} \forall i \leq n \right\} \tag{1}$$

- Set with independent row uncertainties (IRU) : for each i there exists $R_i \subseteq \mathbb{R}_{\geq 0}{}^n$ s.t. :

$$\mathcal{N} = \left\{ (v_1^T, \ldots, v_n^T)^T \mid v_i \in R_i \right\}.$$

Theorem (Blondel & Nesterov '09)

If the IRU set $\mathcal{N}$ is composed of nonnegative matrices, then

$$\rho(\mathcal{N}) = \max\{\rho(A) \mid A \in \mathcal{N}\}.$$

Entropy of tree automata (2)

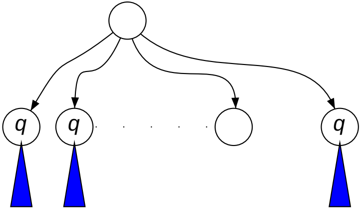
Proposition

For any tree automaton $\mathcal{A}$, $\mathcal{H}(\mathcal{A}) = \log \rho(\mathcal{M}_{\mathcal{A}}) = \max\{\rho(A) \mid A \in \mathcal{M}(\mathcal{A})\}$.

- Remark : there exist trees t for which different transitions might be used at distinct nodes labeled with the same state.
- Hence, $card(t\big|_n)$ might not belong to the set $\left\{\|M_1 \cdot \ldots \cdot M_n\|^{1/n} \mid M_i \in \mathcal{N} \forall i \leq n\right\}$.

Entropy of tree automata (3)

- A Q-tree is isolevel if for all $x_1, x_2 \in \text{supp}(t)$, with $|x_1| = |x_2| = k$ and $t(x_1) = t(x_2)$ we have that $\forall k \in \mathbb{N}, (t(x_1), t(x_1 1), \ldots, t(x_1 k)) = (t(x_2), t(x_2 1), \ldots, t(x_2 k))$.

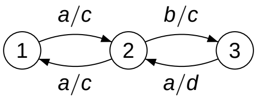


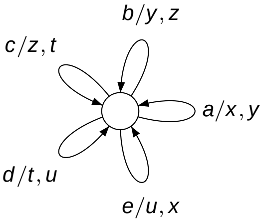
  - An isolevel tree is not a regular tree!

- Proof technique : show that, for each tree t, there exists an isolevel tree t' s.t. $\mathcal{H}(t) \leq \mathcal{H}(t')$.
  - Suppose some t'_n is built upto level n (approximating t' upto level n).
  - For each state q occurring in $(t'_n)\big|_{n+1}$, choose, among all trees with root q, the tree $t'_n[q]$ which contributes the most to $\mathcal{H}(t)$.
  - Build t'_{n+1} by using only $t'_n[q]$.

Generalized zero-error coding

- Synchronous transducer $\mathcal{T} = (Q, \Sigma, \Delta, \theta, q_0)$, $\delta \subseteq Q \times \Sigma \times \Delta \times Q$.
- $L \subseteq \Sigma^n$ is distinguishable if $\forall w_1, w_2 \in L$, $\theta(q_0, w_1) \cap \theta(q_0, w_2) = \emptyset$.
  - Similar definition for $L \subseteq \Sigma^\omega$.
- Examples :





- The capacity of $\mathcal{T}$ is

$$\mathcal{C}(\mathcal{T}) = \limsup_n \left\{ \mathcal{H}(L) \mid L \subseteq \Sigma^n, L \text{ distinguishable} \right\}$$

- The ω-capacity of $\mathcal{T}$ is :

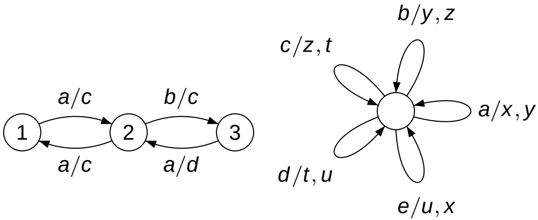
$$\mathcal{C}^\omega(\mathcal{T}) = \sup \left\{ \mathcal{H}(L) \mid L \subseteq \Sigma^\omega, L \text{ is distinguishable} \right\}$$

- The ω-regular capacity of $\mathcal{T}$ is :

$$\mathcal{C}_r^\omega(\mathcal{T}) = \sup \left\{ \mathcal{H}(L) \mid L \subseteq \Sigma^\omega, L \text{ is distinguishable and } \omega\text{-regular} \right\}$$

Generalized zero-error coding

- Synchronous transducer $\mathcal{T} = (Q, \Sigma, \Delta, \theta, q_0)$, $\delta \subseteq Q \times \Sigma \times \Delta \times Q$.
- $L \subseteq \Sigma^n$ is distinguishable if $\forall w_1, w_2 \in L$, $\theta(q_0, w_1) \cap \theta(q_0, w_2) = \emptyset$.
  - Similar definition for $L \subseteq \Sigma^\omega$.
- Examples :



- The capacity of $\mathcal{T}$ is

$$\mathcal{C}(\mathcal{T}) = \limsup_n \{\mathcal{H}(L) \mid L \subseteq \Sigma^n, L \text{ distinguishable}\}$$

- The ω-capacity of $\mathcal{T}$ is :

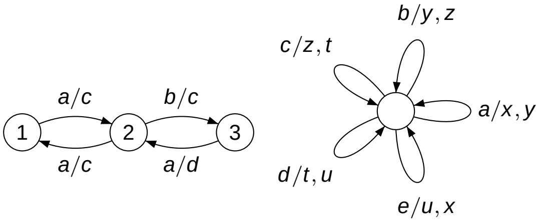
$$\mathcal{C}^\omega(\mathcal{T}) = \sup \{\mathcal{H}(L) \mid L \subseteq \Sigma^\omega, L \text{ is distinguishable}\}$$

- The ω-regular capacity of $\mathcal{T}$ is :

$$\mathcal{C}_r^\omega(\mathcal{T}) = \sup \{\mathcal{H}(L) \mid L \subseteq \Sigma^\omega, L \text{ is distinguishable and } \omega\text{-regular}\}$$

Generalized zero-error coding

- Synchronous transducer $\mathcal{T} = (Q, \Sigma, \Delta, \theta, q_0)$, $\delta \subseteq Q \times \Sigma \times \Delta \times Q$.
- $L \subseteq \Sigma^n$ is distinguishable if $\forall w_1, w_2 \in L$, $\theta(q_0, w_1) \cap \theta(q_0, w_2) = \emptyset$.
  - Similar definition for $L \subseteq \Sigma^\omega$.
- Examples :



- The capacity of $\mathcal{T}$ is

$$\mathcal{C}(\mathcal{T}) = \limsup_n \left\{ \mathcal{H}(L) \mid L \subseteq \Sigma^n, L \text{ distinguishable} \right\}$$

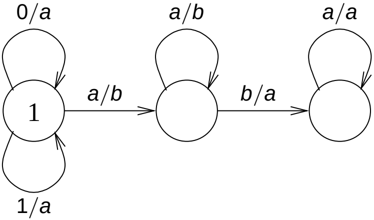
- The ω-capacity of $\mathcal{T}$ is :

$$\mathcal{C}^\omega(\mathcal{T}) = \sup \left\{ \mathcal{H}(L) \mid L \subseteq \Sigma^\omega, L \text{ is distinguishable} \right\}$$

- The ω-regular capacity of $\mathcal{T}$ is :

$$\mathcal{C}^\omega_r(\mathcal{T}) = \sup \left\{ \mathcal{H}(L) \mid L \subseteq \Sigma^\omega, L \text{ is distinguishable and } \omega\text{-regular} \right\}$$

$\mathcal{C}^\omega$ is "unnatural"



- Consider the non-ω-regular language :

$$L_\omega = \{ z a^{\overline{z}_2 + 1} b^\omega \mid z \in (0+1)^* \},$$

  - Here $\overline{z}_2$ is the value of z as an integer written in base 2.
- L_ω is $\mathcal{T}$-distinguishable :
  - $\forall w \in L \; \exists! z \in \mathbb{N}$ with $\mathcal{T}(w) = a^{\lfloor \log_2 z+1 \rfloor} b^z a^\omega$.
- Hence

$$\mathcal{C}^\omega(\mathcal{T}) \geq \mathcal{H}(L_\omega) = 1$$

- But

$$\mathcal{H}(\mathcal{L}_{out}(TTT)) = \mathcal{H}(a^* b b^* a a^\omega) = 0!$$

Relating $\mathcal{C}(\mathcal{T})$ and $\mathcal{C}_r^\omega(\mathcal{T})$

Proposition

$\mathcal{C}_r^\omega(\mathcal{T}) \leq \mathcal{C}(\mathcal{T})$.

- Given $R \subseteq \Sigma^\omega$ distinguishable, any prefix $R[1\ldots n]$ can be "extended" to a distinguishable $R'_n \subseteq \Sigma^{n+N_R}$ for some N_R depending on R.

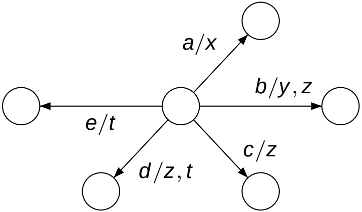
Proposition

If $\mathcal{T}$ is connected and has a synchronizing word then $\mathcal{C}_r^\omega(\mathcal{T}) = \mathcal{C}(\mathcal{T})$.

- Given $L \subseteq \Sigma^n$ with $card(L) \geq \mathcal{C}(\mathcal{T}) - \epsilon$, create a regular tree from L by piling copies of L.
- The extra paths needed to connect two copies of L is bounded by the length of the synchronizing word + the diameter of the automaton.
- The hypothesis on the synchronizing word can be weakened to requesting that, for any set of states $S \subseteq Q$ in the *determinization of* $\mathcal{T}_{in}$, there exists a state q and a word w such that $r \xrightarrow{w} q$ for any $r \in S$.
  - Property which trivially holds when $\mathcal{T}_{in}$ is deterministic.

Tree automata from (special) channels

- Assume $\mathcal{T}_{in} = (Q, \Sigma, \delta_{in}, q_0)$ is deterministic.
- Construct a tree automaton $\mathcal{A}_{\mathcal{T}}$:
  - At each state choose distinguishable sets of outgoing transitions :

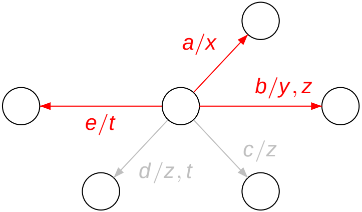


  - $(q, S) \in \theta$ if for each $r_1, r_2 \in S$, if $q \xrightarrow{a/x} r_1$ $q \xrightarrow{b/y} r_2$ for $a \neq b$ and $x \neq y$ then $r_1 = r_2$.
- Then

$$\rho(\mathcal{A}) \leq \mathcal{C}(\mathcal{T})$$

Tree automata from (special) channels

- Assume $\mathcal{T}_{in} = (Q, \Sigma, \delta_{in}, q_0)$ is deterministic.
- Construct a tree automaton $\mathcal{A}_{\mathcal{T}}$:
  - At each state choose distinguishable sets of outgoing transitions :

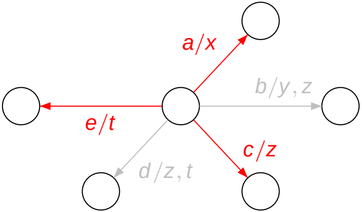


  - $(q, S) \in \theta$ if for each $r_1, r_2 \in S$, if $q \xrightarrow{a/x} r_1$ $q \xrightarrow{b/y} r_2$ for $a \neq b$ and $x \neq y$ then $r_1 = r_2$.
- Then

$$\rho(\mathcal{A}) \leq \mathcal{C}(\mathcal{T})$$

Tree automata from (special) channels

- Assume $\mathcal{T}_{in} = (Q, \Sigma, \delta_{in}, q_0)$ is deterministic.
- Construct a tree automaton $\mathcal{A}_{\mathcal{T}}$:
  - At each state choose distinguishable sets of outgoing transitions :



  - $(q, S) \in \theta$ if for each $r_1, r_2 \in S$, if $q \xrightarrow{a/x} r_1$ $q \xrightarrow{b/y} r_2$ for $a \neq b$ and $x \neq y$ then $r_1 = r_2$.
- Then

$$\rho(\mathcal{A}) \leq \mathcal{C}(\mathcal{T})$$

Tree automata from channels (2)

- Assume $\mathcal{T}_{in} = (Q, \Sigma, \delta_{in}, q_0)$ is deterministic.
- For each n, construct the tree automaton $\mathcal{A}^n_{\mathcal{T}}$:
  - At each state choose distinguishable sets of runs of length n.
  - $(q, S) \in \theta$ if for each $r_1, r_2 \in S$, if $q \xrightarrow{w_1/x_1} r_1$ and $q \xrightarrow{w_2/x_2} r_2$ for $w_1 \neq w_2$, $w_1, w_2 \in \Sigma^n$, and $x \neq y$ then $r_1 = r_2$.

Proposition

$\lim_{n\to\infty} \frac{1}{n} \log_2 \rho(\mathcal{A}^n_{\mathcal{T}}) = \mathcal{C}(\mathcal{T})$ when $\mathcal{T}_{in}$ is deterministic.

Open problems

- Does always $\mathcal{C}_r^\omega(\mathcal{T}) = \mathcal{C}(\mathcal{T})$?
- Characterize transducers for which $\lim_{n\to\infty} \frac{1}{n} \log_2 \rho(\mathcal{A}_\mathcal{T}^n) = \mathcal{C}(\mathcal{T})$.
- Explore the possibility to express $\mathcal{H}(t) \geq \mathcal{H}(t')$ using some enrichment of MSO with special predicates.