Introduction to Symbolic Dynamics

Part 4: Entropy

Silvio Capobianco

Institute of Cybernetics at TUT

May 12, 2010

Revised: May 12, 2010

Overview

- Constructions and algorithms on sofic shifts.
- Entropy of a shift subspace.
- Computing entropy via Perron-Frobenius theory.

Sofic shifts

Path labelings

Let $\mathcal{G} = (G, \mathcal{L})$ be an A-labeled graph.

- The labeling of a path $\pi = e_1 \ldots e_m$ on G is the sequence $\mathcal{L}(\pi) = \mathcal{L}(e_1) \ldots \mathcal{L}(e_m)$.
- The labeling of a bi-infinite path $\xi \in \mathcal{E}(G)^{\mathbb{Z}}$ is the sequence $x = \mathcal{L}(\xi) \in A^{\mathbb{Z}}$ s.t. $x_i = \mathcal{L}(\xi_i)$ for every $i \in \mathbb{Z}$.
- We put

$$\mathsf{X}_{\mathcal{G}} = \left\{ x \in A^{\mathbb{Z}} \mid \exists \xi \in \mathsf{X}_G \mid x = \mathcal{L}(\xi) \right\}$$

Definition

- $X \subseteq A^{\mathbb{Z}}$ is a sofic shift if $X = \mathsf{X}_{\mathcal{G}}$ for some A-labeled graph $\mathcal{G}$.
- In this case, $\mathcal{G}$ is a presentation of X.

Special kinds of presentations

A labeled graph $\mathcal{G} = (G, \mathcal{L})$ is:

- right-resolving if initial state and label determine edge
- follower-separated if differen states have different follower sets

Fischer's theorem

Two minimal right-resolving presentations of an irreducible sofic shifts are isomorphic.

Unions

Union of two graphs

Let $\mathcal{G}_1 = (G_1, \mathcal{L}_1)$ and $\mathcal{G}_2 = (G_1, \mathcal{L}_2)$ be labeled graphs.

- Set $\mathcal{V}(G) = \mathcal{V}(G_1) \sqcup \mathcal{V}(G_2)$.
- Set $\mathcal{E}(G) = \mathcal{E}(G_1) \sqcup \mathcal{E}(G_2)$.
- Set $\mathcal{L}(e) = \mathcal{L}_i(e)$ if $e \in \mathcal{E}(G_i)$

Then $\mathcal{G} = (G, \mathcal{L}) = \mathcal{G}_1 \cup \mathcal{G}_2$ is the union of $\mathcal{G}_1$ and $\mathcal{G}_2$.

Union of two sofic shifts is sofic

$\mathcal{G}_1 \cup \mathcal{G}_2$ is a presentation of $\mathsf{X}_{\mathcal{G}_1} \cup \mathsf{X}_{\mathcal{G}_2}$.

Products

Product of two graphs

Let $\mathcal{G}_1 = (G_1, \mathcal{L}_1)$ and $\mathcal{G}_2 = (G_1, \mathcal{L}_2)$ be labeled graphs.

- Set $\mathcal{V}(G) = \mathcal{V}(G_1) \times \mathcal{V}(G_2)$.
- Set $\mathcal{E}(G) = \mathcal{E}(G_1) \times \mathcal{E}(G_2)$.
- Set $\mathcal{L}(e) = \mathcal{L}(e_1, e_2) = (\mathcal{L}_1(e_1), \mathcal{L}_2(e_2))$.

Then $\mathcal{G} = (G, \mathcal{L}) = \mathcal{G}_1 \times \mathcal{G}_2$ is the product of $\mathcal{G}_1$ and $\mathcal{G}_2$.

Product of two sofic shifts is sofic

$\mathcal{G}_1 \times \mathcal{G}_2$ is a presentation of $X_{\mathcal{G}_1} \times X_{\mathcal{G}_2}$.

Label products

Label product of two graphs

Let $\mathcal{G}_1 = (G_1, \mathcal{L}_1)$ and $\mathcal{G}_2 = (G_1, \mathcal{L}_2)$ be labeled graphs.

- Set $\mathcal{V}(G) = \mathcal{V}(G_1) \times \mathcal{V}(G_2)$.
- Set $\mathcal{E}(G) = \{(e_1, e_2) \in \mathcal{E}(G_1) \times \mathcal{E}(G_2) \mid \mathcal{L}_1(e_1) = \mathcal{L}_2(e_2)\}$.
- Set $\mathcal{L}(e) = \mathcal{L}(e_1, e_2) = \mathcal{L}_1(e_1) = \mathcal{L}_2(e_2)$.

Then $\mathcal{G} = (G, \mathcal{L}) = \mathcal{G}_1 * \mathcal{G}_2$ is the label product of $\mathcal{G}_1$ and $\mathcal{G}_2$.

Intersection of two sofic shifts is sofic

$\mathcal{G}_1 * \mathcal{G}_2$ is a presentation of $\mathsf{X}_{\mathcal{G}_1} \cap \mathsf{X}_{\mathcal{G}_2}$.

And it isn't over here...

If $\mathcal{G}_1$ and $\mathcal{G}_2$ are right-resolving, then $\mathcal{G}_1 * \mathcal{G}_2$ is right-resolving.

Equality of sofic shifts

The problem

Given $\mathcal{G}_1$ and $\mathcal{G}_2$, determine whether $\mathsf{X}_{\mathcal{G}_1} = \mathsf{X}_{\mathcal{G}_2}$.

The idea

Express equality of sofic shifts through the constructions seen before.

An useful lemma

- Let $G = (\mathcal{V}, \mathcal{E})$ be a graph with r states.
- Let $\mathcal{S} \subseteq \mathcal{V}$ contain s states.
- For $I \in \mathcal{V} \setminus \mathcal{S}$ let $U_I = \{\pi \text{ path on } G \mid \mathrm{i}(\pi) = I, \mathrm{t}(\pi) \in \mathcal{S}\}$.
- If U_I is nonempty then $\min\{|\pi| \mid \pi \in U_I\} \leq r - s$.
- Thus, there is a path from $I \notin \mathcal{S}$ to $J \in \mathcal{S}$ iff $B_{I,J} > 0$, where $B = \sum_{i=1}^{r-s} A^i$.

Equality of sofic shifts is decidable

The idea

Given $\mathcal{G}_1$ and $\mathcal{G}_2$, construct $\widehat{\mathcal{G}}$ s.t. TFAE:

1. There is a word in $\mathcal{B}(\mathsf{X}_{\mathcal{G}_i}) \setminus \mathcal{B}(\mathsf{X}_{\mathcal{G}_j})$.
2. There is a path in $\widehat{\mathcal{G}}$ from some state $\mathcal{I}$ to some set $\mathcal{S}_i$.

The algorithm

1. Let $\mathcal{G}_i'$ be $\mathcal{G}_i$ plus a sink K_i: If there is no edge from I labeled a, make an edge from I to K_i labeled a; Add all self-loops to K_i.
2. Let $\widehat{\mathcal{G}_i}$ be the subset graph of $\mathcal{G}_i'$. Let $\mathcal{K}_i = \{K_i\}$.
3. Let $\widehat{\mathcal{G}} = \widehat{\mathcal{G}_1} * \widehat{\mathcal{G}_2}$. Set $\mathcal{I} = (\mathcal{V}_1, \mathcal{V}_2)$.
4. Set $\mathcal{S}_1 = \{(\mathcal{J}, \mathcal{K}_2) \mid \mathcal{J} \neq \mathcal{K}_1\}$ and $\mathcal{S}_2 = \{(\mathcal{K}_1, \mathcal{J}) \mid \mathcal{J} \neq \mathcal{K}_2\}$

Cost of the algorithm

If $\mathcal{G}_i$ has r_i states...

...then $\widehat{\mathcal{G}}$ has $(2^{r_1+1} - 1) \cdot (2^{r_2+1} - 1)$.

Could one do better?

- In general, no.
- But maybe, in special cases...

A hint from Fischer's theorem

- Suppose $\mathcal{G}_1$ and $\mathcal{G}_2$ are irreducible and right-resolving.
- Let $\mathcal{H}_i$ be the minimal right-resolving presentation of $\mathcal{X}_{\mathcal{G}_i}$.
- Then $\mathsf{X}_{\mathcal{G}_1} = \mathsf{X}_{\mathcal{G}_2}$. if and only if $\mathcal{H}_1 \cong \mathcal{H}_2$.

Constructing the minimal right-resolving presentation

The idea

- Start from an irreducible right-resolving presentation.
- Its merged graph is the minimal right-resolving presentation.

Deciding equality of follower sets

1. Let $\mathcal{G}'$ be $\mathcal{G}$ with a sink K, as before.
2. Set $\widehat{\mathcal{G}} = \mathcal{G}' * \mathcal{G}'$, $\mathcal{I} = \mathcal{V} \times \mathcal{V}$, $\mathcal{S} = (\mathcal{V} \times \{K\}) \cup (\{K\} \times \mathcal{V})$.
3. Let I and J be two distinct nodes in $\mathcal{G}$. TFAE.
   - $F_{\mathcal{G}}(I) \neq F_{\mathcal{G}}(J)$.
   - There is a path from (I, J) to $\mathcal{S}$ in $\widehat{\mathcal{G}}$.

Determining finiteness of type of sofic shifts

Theorem A

- Let $\mathcal{G}$ be a right-resolving labeled graph.
- Suppose that every $w \in \mathcal{B}_N(\mathsf{X}_{\mathcal{G}})$ is synchronizing for $\mathcal{G}$.
- Then $\mathsf{X}_{\mathcal{G}}$ is an N-step SFT.

Theorem B

- Let X be an irreducible sofic shift.
- And let $\mathcal{G} = (G, \mathcal{L})$ be its minimal right-resolving presentation.
- Suppose that X is an N-step SFT.
- Then:
  - Every $w \in \mathcal{B}_N(\mathsf{X}_{\mathcal{G}})$ is synchronizing for $\mathcal{G}$.
  - $\mathcal{L}_\infty$ is a conjugacy.
  - If G has r states then X is $(r^2 - r)$-step.

Proof of Theorems A and B

Proof of Theorem A

- Suppose $uw, wv \in \mathcal{B}(\mathsf{X}_{\mathcal{G}})$ with $w \geq N$—then w is synchronizing.
- If $uw = \mathcal{L}(\rho\pi)$ and $wu = \mathcal{L}(\tau\sigma)$, then $\mathrm{t}(\pi) = \mathrm{t}(\tau)$.
- Then $uwv = \mathcal{L}(\rho\pi\sigma) \in \mathcal{B}(\mathsf{X}_{\mathcal{L}})$.

Proof of Theorem B

- Suppose $|w| = N$ and $w = \mathcal{L}(\pi) = \mathcal{L}(\tau)$ with $\mathrm{t}(\pi) \neq \mathrm{t}(\tau)$...
  - Let $v \in F_{\mathcal{G}}(\mathrm{t}(\pi)) \setminus F_{\mathcal{G}}(\mathrm{t}(\tau))$, u synchronizing word focusing on $\mathrm{i}(\tau)$.
  - Then $uw, wv \in \mathcal{B}(X)$ but $uwv \notin \mathcal{B}(X)$, against X being N-step.
- If $x = \mathcal{L}_\infty(y) = \mathcal{L}_\infty(z)$, then $y_{[i,\infty)} = z_{[i,\infty)}$ because $\mathcal{L}(y_{[i-N,i-1]}) = \mathcal{L}(z_{[i-N,i-1]})$ is synchronizing and $\mathcal{G}$ is right-resolving.
- The graph $\mathcal{G} * \mathcal{G}$ minus diagonal vertices checks precisely non-synchronizing words and has $r^2 - r$ states.

Entropy

Definition

The entropy of a nonempty shift X is

$$h(X) = \lim_{n\to\infty} \frac{1}{n} \log |\mathcal{B}_n(X)| = \inf_{n\geq 1} \frac{1}{n} \log |\mathcal{B}_n(X)|$$

The limit above exists and the equality holds because for every $m, n \geq 1$

$$|\mathcal{B}_{m+n}(X)| \leq |\mathcal{B}_m(X)| \cdot |\mathcal{B}_n(X)|$$

If $X = \emptyset$ we put $h(X) = -\infty$.

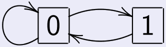
Quick examples

- If X is a full shift on an alphabet of r elements then $h(X) = \log r$.
- If G is a graph on k nodes with r outgoing edges per node then $h(\mathrm{X}_G) = \log r$.

The entropy of the golden mean shift

The idea for the computation

- Consider the even shift as the vertex shift of



- For $n \geq 2$ there is a one-to-one correspondence between $\mathcal{B}_n(\widehat{\mathsf{X}}_G)$ and $\mathcal{B}_{n-1}(\mathsf{X}_G)$.
- We can compute the size of this through the adjacency matrix

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

because

$$|\mathcal{B}_m(\mathsf{X}_G)| = (A^m)_{0,0} + (A^m)_{0,1} + (A^m)_{1,0} + (A^m)_{1,1}$$

The entropy of the golden mean shift (cont.)

Eigenvalues and eigenvectors

- The characteristic polynomial of A is

$$\chi_A(t) = t^2 - t - 1$$

  which has solutions

$$\lambda = \frac{1+\sqrt{5}}{2} \; ; \; \mu = \frac{1-\sqrt{5}}{2}$$

  λ is known as the golden mean.

- Corresponding eigenvectors of A are

$$\mathbf{v}_\lambda = \begin{pmatrix} \lambda \\ 1 \end{pmatrix} \; ; \; \mathbf{v}_\mu = \begin{pmatrix} \mu \\ 1 \end{pmatrix}$$

The entropy of the golden mean shift (end)

Diagonalizing

- Let $P = \begin{pmatrix} \lambda & \mu \\ 1 & 1 \end{pmatrix}$. Then

$$A^m = P \begin{pmatrix} \lambda^m & 0 \\ 0 & \mu^m \end{pmatrix} P^{-1} = \frac{1}{\sqrt{5}} \begin{pmatrix} \lambda^{m+1} - \mu^{m+1} & \lambda^m - \mu^m \\ \lambda^m - \mu^m & \lambda^{m-1} - \mu^{m-1} \end{pmatrix}$$

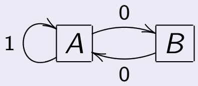
- But $\lambda^{m+2} = \lambda^{m+1} + \lambda^m$ because $\lambda^2 = \lambda + 1$, and similar with μ.
- Hence $|\mathcal{B}_n(\widehat{X}_G)| = \frac{1}{\sqrt{5}}(\lambda^{n+2} - \mu^{n+2})$, from which

$$\begin{aligned} h(\widehat{X}_G) &= \lim_{n\to\infty} \frac{1}{n} \log \left(\frac{1}{\sqrt{5}} \lambda^{n+2} \left(1 - \left(\frac{\mu}{\lambda} \right)^{n+2} \right) \right) \\ &= \log \lambda \end{aligned}$$

The entropy of the even shift

The idea for the computation

- Consider the even shift as presented by



- Each word with a 1 has one presentation. 0^n has two presentations.
- Then, $|\mathcal{B}_n(\mathsf{X}_{\mathcal{G}})| = |\mathcal{B}_n(\mathsf{X}_G)| - 1$.
- Then clearly $h(\mathrm{even\,shift}) = h(\mathrm{golden\,mean\,shift}) = \log \lambda$.

Entropy and sliding block codes

Theorem

If Y is a factor of X then $h(Y) \leq h(X)$.

Consequences

- Entropy is a shift invariant, *i.e.*, conjugate shifts have same entropy. In particular, $h(X^{[N]}) = h(X)$.
- If $\mathcal{G} = (G, \mathcal{L})$ then $h(\mathsf{X}_{\mathcal{G}}) \leq h(\mathsf{X}_G)$.
- Two full shifts are conjugate iff the size of their alphabets is the same.
- The golden mean shift is not conjugate to any full shift.
- If Y embeds into X then $h(Y) \leq h(X)$.

Reason why the theorem holds

If $\Phi_{\infty}^{[-m,\alpha]} : X \to Y$ is a factor code, then $|\mathcal{B}_n(Y)| \leq |\mathcal{B}_{m+n+\alpha}(X)|$.

Entropy and labeled graphs

Theorem

Let $\mathcal{G} = (G, \mathcal{L})$ be a right-resolving labeled graph.
Then $h(\mathsf{X}_{\mathcal{G}}) = h(\mathsf{X}_G)$.

Reason why

- Suppose G has k states.
- Since $\mathcal{G}$ is right-resolving, there can be at most k paths representing each $w \in \mathcal{B}(\mathsf{X}_{\mathcal{G}})$.
- Thus, $|\mathcal{B}_n(\mathsf{X}_G)| \leq k \cdot |\mathcal{B}_n(\mathsf{X}_{\mathcal{G}})|$.

Estimates on $|\mathcal{B}_n(\mathsf{X}_G)|$ for "good" $A(G)$

Let G be a graph with at least one edge and A its adjacency matrix.

The key hypothesis

Suppose A has a positive eigenvector $\mathbf{v}$.

A long series of consequences

- The corresponding eigenvalue λ is positive.
- If $m = \min_i v_i$ and $M = \max_i v_i$, then

$$\frac{m}{M}\lambda^n \leq \sum_{I,J=1}^{r} (A^n)_{I,J} \leq \frac{rM}{m}\lambda^n ,$$

  which implies $h(\mathsf{X}_G) = \log \lambda$.
- λ is the only eigenvalue of A corresponding to a positive eigenvector.
- If μ is any other eigenvalue for A, it can be shown that $|\mu| \leq \lambda$.

The Perron-Frobenius theorem

Let A be a nonnegative irreducible nonzero matrix.

1. A has a positive eigenvector $\mathbf{v}_A$.
2. The eigenvalue λ_A corresponding to $\mathbf{v}_A$ is positive.
3. λ_A is algebraically—and geometrically—simple, *i.e.*,
    - $\det(tI - A) = (t - \lambda_A)p(t)$ with $p(\lambda_A) \neq 0$, and
    - $\dim\{\mathbf{v} \mid A\mathbf{v} = \lambda_A \mathbf{v}\} = 1$.
4. If μ is another eigenvalue of A then $|\mu| \leq \lambda_A$.
5. Any positive eigenvector of A is a positive multiple of $\mathbf{v}_A$.

The value λ_A is called the Perron eigenvalue of A

Computing entropy with Perron-Frobenius theorem

Entropy of an irreducible edge shift

If G is an irreducible graph then $h(\mathsf{X}_G) = \log \lambda_{A(G)}$.

Entropy of an irreducible SFT

If X is an irreducible M-step SFT and G is the essential graph s.t. $X^{[M+1]} = \mathsf{X}_G$, then $h(X) = \log \lambda_{A(G)}$.

Entropy of an irreducible sofic shift

If X is an irreducible sofic shift and $\mathcal{G} = (G, \mathcal{L})$ is an irreducible right-resolving presentation of X, then $h(X) = \log \lambda_{A(G)}$.

Entropy and periodic points

Two simple estimates

- Let $p_n(X)$ be the number of points of X with period n.
- Let $q_n(X)$ be the number of points of X with minimum period n.
- Clearly,

$$h(X) \geq \limsup_n \frac{1}{n} p_n(X) \geq \limsup_n \frac{1}{n} q_n(X)$$

Sofic shifts and periodic points

If X is an irreducible sofic shift then

$$\begin{aligned} h(X) &= \limsup_n \frac{1}{n} p_n(X) \\ &= \limsup_n \frac{1}{n} q_n(X) \end{aligned}$$

Proof of the previous theorem

If X is an M-step SFT

- Let G be an irreducible graph s.t. $X \cong \mathsf{X}_G$. Then $p_n(X) = p_n(\mathsf{X}_G)$
- Let N be the maximum length of a shortest path from any two states.
- For every $w \in \mathcal{B}(\mathsf{X}_G)$ there is $u \in \bigcup_{i=0}^{N} \mathcal{B}_i(\mathsf{X}_G)$ s.t. $(wu)^\infty \in \mathsf{X}_G$.
- Thus, $|\mathcal{B}_n(\mathsf{X}_G)| \leq \sum_{i=0}^{N} p_{n+i}(\mathsf{X}_G) \leq (N+1) p_{n+i(n)}(X)$ for some $i(n) \in \{0, \dots, N\}$.

If X is sofic

- Let $\mathcal{G} = (G, \mathcal{L})$ be an irreducible right-resolving presentation of X. Then $h(X) = h(\mathsf{X}_G)$.
- If G has r states, then every labeled path on $\mathcal{G}$ has at most r presentations on G (because of right-resolvingness).
- Then $p_n(X) \geq \frac{1}{r} p_n(\mathsf{X}_G)$ because $\mathcal{L}_\infty$ preserves periods.

. . . but what for reducible shifts?

The idea

- Identify the irreducible components.
- Apply the theory to those.
- Get information on the whole graph.

The procedure

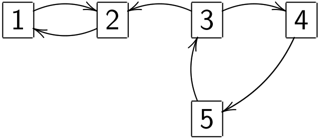
Given G, construct H as follows:

- Nodes in H are irreducible components in G.
- There is an edge from I to J in H iff J is reachable from I in G.
- Order of nodes is such that if I is reachable from J then $I < J$.

This construction determines a re-ordering of the rows and columns of the adjacency matrix of G.
The new matrix is block lower triangular.

Example

Consider the graph



The irreducible components are $\{1, 2\}$ and $\{3, 4, 5\}$—already sorted.
The adjacency matrix is

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix} = \begin{pmatrix} A_1 & 0 \\ B & A_2 \end{pmatrix}$$

Perron-Frobenius theory for reducible matrices

Definition

- Let A be a nonnegative, nonzero matrix
- Suppose A is in block lower triangular form

$$A = \begin{pmatrix} A_1 & 0 & \dots & 0 \\ * & A_2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ * & * & \dots & A_k \end{pmatrix}$$

  with each A_i irreducible.
- The Perron eigenvalue of A is $\lambda_A = \max_{1 \leq i \leq k} \lambda_{A_i}$

Motivation

λ_A is the largest eigenvalue of A.

Entropy of graph shifts

Theorem

For any graph G, $h(\mathsf{X}_G) = \log \lambda_{A(G)}$.

Corollary

For any right-resolving labeled graph $\mathcal{G} = (G, \mathcal{L})$, $h(\mathsf{X}_{\mathcal{G}}) = \log \lambda_{A(G)}$.

Idea of the proof

- Each path is a chain of paths on irreducible components linked by single edges.
- Single edges can occur in at most n places, and get at most M values.
- On the j-th component, $|\mathcal{B}_{n(j)}(\mathsf{X}_{G_j})| \leq D \cdot \lambda_A^{n(j)}$ for a suitable D.

Approximating entropy

Theorem

If $\{X_k\}_{k \geq 1}$ is a monotone non-increasing family of subshifts, then

$$\lim_{k \to \infty} h(X_k) = h\left(\bigcap_{k=1}^{\infty} X_k\right)$$

Reason why

- Put $X = \bigcap_{k=1}^{\infty} X_k$.
- For every n exists $k(n)$ s.t. $\mathcal{B}_n(X_k) = \mathcal{B}_n(X)$ for every $k \geq k(n)$. Otherwise, find x in all X_k's but not in X via a diagonal argument.
- Thus, if $\frac{1}{n} \log |\mathcal{B}_n(X)|$ does not exceed $h(X) + \varepsilon$, neither does $h(X_k) = \inf_{n \geq 1} \frac{1}{n} \log |\mathcal{B}_n(X_k)|$ for $k \geq k(n)$.

Approximating entropy (cont.)

Problems with previous approach

- Computing $h(X_k)$ can become difficult for high k, especially if the X_k's are edge shifts on graphs of increasing size.
- It is not clear which k's provide the desired approximation!

Inside sofic approximation

- Idea: find an irreducible sofic $Y \subseteq X$.
- Advantage: $h(Y) \leq h(X) \leq h(X_k)$, so check if $h(X_k) - h(Y) < \varepsilon$.
- Disadvantage: it is not clear how Y should be built.
- In fact, there are shift spaces with no nonempty sofic subshift!

Soon on these screens...

- Cyclic structure of irreducible matrices
- The road problem
- The finite-state coding theorem

Thank you for attention!