LECTURE 2 - SOME ASPECTS OF ENTROPY VIA EXERCISES

NISHANT CHANDGOTIA

1. Some simple computations

It is not possible to give a comprehensive introduction to entropy in a period of forty five minutes. We will give a very specialised introduction for our purposes by means of certain exercises.

Recall that every shift space $X \subset \mathcal{A}^{\mathbb{Z}^d}$ is obtained as the set of configurations avoiding patterns from a forbidden list $\mathcal{F}$. $B_n := \{1, 2, \ldots, n\}^d$ is a box of size n. The language allowed on the shape B_n is given by

$$\mathcal{L}(X, B_n) := \{x|_{B_n} \ : \ x \in X\}.$$

Then the entropy is defined as

$$h_{\text{top}}(X) := \lim_{n\to\infty} \frac{1}{|B_n|} \log(|\mathcal{L}(X, B_n)|).$$

The hard-core shift is the subshift with alphabet $\{0, 1\}$ where adjacent 1's are disallowed. The even shift is the subshift with alphabet $\{0, 1\}$ where the gap between successive 1's is even. The space of proper k-colorings is the subshift with alphabet $\{0, 1, \ldots, k-1\}$ where adjacent symbols are forced to be distinct.

For the following question let us remind ourselves of the following consequence of the Perron-Frobenius theorem.

Theorem 1. [10, Proposition 4.2.1] *Let A be an integer matrix with non-negative entries. Then A has a positive eigen-value $\lambda > 0$ such that there exist $c, d > 0$*

$$c\lambda^n \leq \vec{1}^t A^n \vec{1} \leq d\lambda^n.$$

for all $n \in \mathbb{N}$.

Question 1. *Compute the topological entropy of the hard-core shift. Following the idea in the previous lecture (Figure 1) show that the even shift is the factor of the hard-core in $d = 1$. Can you use this to find the topological entropy of the even shift? (Hint: To find the entropy of the hard-core shift try to form a recurrence relation for $\mathcal{L}(X, B_n)$ where X is the hard-core shift and then use Theorem 1.) Prove that the entropy of the space of proper k-colorings is $\log(k-1)$.*

Looking at this you might have a sense of how algebraic integers turn up in the context of one dimensional SFTs. For more about this in the case of $d = 1$ look at [10, Chapter 4]. In general computation of these constants can be a very difficult task in higher dimensions. In $d = 2$ the hard-core shift is one of the few where algorithms are known for approximation within polynomial time [13]. In the case of 3-colorings the entropy is explicitly known due to Lieb [9] (also look at

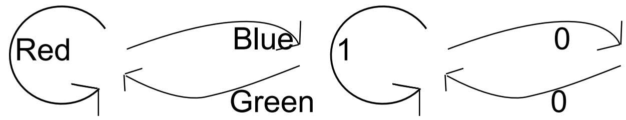

FIGURE 1. An SFT and the even shift as its factor

[2]). In general approximating or calculating entropy in higher dimensions even for simple seeming models can be very difficult.

2. Entropy and continuous maps

Recall that a factor map $\phi : X \to Y$ is a continuous surjective map which is equivariant, meaning, it commutes with the shift map, that is, $\sigma_Y \circ \phi = \phi \circ \sigma_X$.

Question 2. *Prove that if Y is a factor of X then $h_{top}(Y) \leq h_{top}(X)$.*

Let me introduce a theorem to you which would help. By expanding the alphabet if necessary we can assume that $X, Y \subset \mathcal{A}^{\mathbb{Z}^d}$. A sliding block map is a map $\phi : X \to Y$ with the following structure:

There exists $B \subset \mathbb{Z}^d$ and a map $\Phi : \mathcal{L}(X, B) \to \mathcal{A}$ such that $\phi(x)_{\vec{i}} := \Phi(\sigma^{\vec{i}}(x)|_B)$.

Theorem 2 (Curtis-Hedlund-Lyndon)**.** [10, Section 1.5] *A map $\phi : X \to Y$ is equivariant and continuous if and only if ϕ is a sliding block code.*

As a further comment suppose $\phi : X \to Y$ is injective as well. Since the image of compact sets is compact it follows that the image of closed sets under ϕ are closed and hence the image of open sets are open. Thus ϕ is invertible and one gets as a corollary that $h_{\text{top}}(Y) = h_{\text{top}}(X)$. Such maps are called a conjugacy.

3. Nearest neighbour SFTs

An SFT X is called a *nearest neighbour SFT* if $X = X_{\mathcal{F}}$ for some set of patterns $\mathcal{F}$ on the shapes $\{\vec{0}, \vec{e}_i\}$ for $1 \leq i \leq d$. Clearly the examples that we have discussed until now (hard core shift, space of proper colourings, dimer tilings) are all nearest neighbour SFTs.

Question 3. *Prove that shift spaces conjugate to SFTs are still SFTs and that every SFT is conjugate to a nearest neighbour SFT. (Hint: Use Curtis-Hedlund-Lyndon theorem)*

4. Measure theoretic entropy

For this section let us remind ourselves of Jensen's inequality.

Theorem 3 (Jensen's Inequality)**.** *Let $(\Omega, \mu, \mathcal{B})$ be a probability space and $f \in L^1(\mu)$. Then for all convex functions $\phi : \mathbb{R} \to \mathbb{R}$,*

$$\phi\left(\int_\Omega f(\omega)d\mu(\omega)\right) \leq \int_\Omega \phi \circ f(\omega)d\mu(\omega)$$

where equality occurs if ϕ is linear on the image of f or f is a constant function.

The *Shannon entropy* of probability measure μ on a countable set N is defined as

$$H(\mu) := \sum_{n\in N} -\mu(n)\log(\mu(n))$$

where $0.log(0) = 0$.

Entropy takes value in $[0,\infty]$. Here are some elementary exercises to familiarise yourself with entropy.

Question 4. *(1) $H(\mu) = 0$ if and only if μ is deterministic, that is, $\mu(x) = 1$ for some $x \in N$.*
(2) If μ is a probability measure on a finite set N then prove that the entropy is maximised for the uniform probability measure. (Hint: $x \to -\log(x)$ is a convex function)

5. Measures on shift spaces

Given a graph G and a set of vertices $B \subset G$, let

$$\partial B := \{v \in G \setminus B \ : \ v \text{ is adjacent to a vertex in } B\}.$$

Let μ be a probability measure on $\mathcal{A}^G$. By the *support of the measure* μ we mean the smallest closed set $X \subset \mathcal{A}^G$ for which $\mu(X) = 1$. In other words

$$supp(\mu) = \mathcal{A}^G \setminus (\cup_\star [b]_C)$$

where $\star$ is the set of all cylinder set $[b]_C := \{x \in \mathcal{A}^G \ : \ x|_C = b\}$ for which $\mu([b]_C) = 0$.

A shift-invariant probability measure μ on a shift space X is a probability measure which is left invariant under the shift action. More concretely,

$$\mu([b]_C) := \mu(\sigma^{\vec{i}}([b]_C)) \text{ for all } \vec{i} \in \mathbb{Z}^d.$$

Clearly $supp(\mu)$ is also a shift space since if b is a pattern which does not appear in $supp(\mu)$ then so doesn't $\sigma^{\vec{i}}(b)$ for all $\vec{i} \in \mathbb{Z}^d$.

We will use the notation for a measure μ on X and measurable sets $D, E \subset X$; $\mu(E) > 0$, the notation $\mu(D \ : \ E) := \mu(D \cap E)/\mu(E)$.

Question 5. *Let X be a nearest neighbour SFT. Prove that $H(\mu)$ is maximised over all possible probability measures μ on $\mathcal{L}(X, B_n)$ exactly when μ is a uniform Gibbs measure, meaning, that for all finite sets $B \subset B_n$ and $a \in supp(\mu)$ and $b \in \mathcal{L}(X, B_n)$,*

$$\mu([b]_B \ : \ [a]_{B_n \setminus B}) = \mu([b]_B \ : \ [a]_{\partial B \cap B_n})$$

and is uniform on $b \in \mathcal{L}(X, B \cup \partial B)$ such that $b|_{\partial B} = a|_{\partial B}$.

6. Existence of measures of maximal entropy for shift spaces

Given a measure μ on a shift space, we can restrict it to $\mathcal{L}(X, B_n)$ in the following natural way,

$$\mu_{B_n}(a) := \mu([a]_{B_n}).$$

We will need to recall the Banach-Alaoglu theorem.

Theorem 4 (Banach-Alaoglu theorem)**.** *Given a normed space H the unit ball in its dual H^* is compact under the weak-star topology.*

Given a shift space X, let $\mathcal{P}(X)$ denote the space of all shift-invariant probability measures on X. Under the weak topology $\mathcal{P}(X)$ is a compact space (by the Banach-Alaoglu theorem but in our particular case it is just a diagonalisation argument). The *measure-theoretic entropy* of μ is given by

$$h_\mu := \lim_{n\to\infty} \frac{1}{|B_n|} H(\mu_{B_n}).$$

Again by subadditivity it can be proven that the limit exists and that

$$h_\mu = \inf \frac{1}{|B_n|} H(\mu_{B_n}).$$

Question 6. *(1) Prove that $\mu \to h_\mu$ is upper-semi continuous and that there is a measure $\mu \in \mathcal{P}(X)$ which maximises measure-theoretic entropy.*

(2) Construct an example to exhibit the fact that $\mu \to h_\mu$ need not always be continuous. (Hint: Let μ be the iid measure on $\{0,1\}^{\mathbb{Z}}$ and find a sequence of measures suppported on periodic points which converge to it.)

7. Some final notes

In general measure-theoretic entropy is not defined in this way. Given a $\mathbb{Z}^d$-action on a probability space (X, μ, B), the measures theoretic entropy (as first formulated by Kolmogorov and Sinai in 1958) is defined as the maximum of the measure theoretic entropy of random fields (measures on shift spaces) obtained by taking finite measurable partitions of X. That this general definition coincides with ours is a result of their theorem which says that if the partition generates the sigma-algebra of the space, then the entropy of the $\mathbb{Z}^d$- action is the same as the entropy of the resulting process. More details can be found in [8, Chapter 3].

An important result that we will need is the following:

Theorem 5 (Variational Principle)**.** *For a shift space X we have that*

$$\sup_{\mu\in\mathcal{P}(X)} h_\mu = h_{top}(X).$$

A beautiful short proof can be found in [11]. Again the general result extends to actions of $\mathbb{Z}^d$ (and beyond, to amenable groups) by homeomorphisms on compact metric spaces. We remark though that in this generality, the measure theoretic entropy is not necessarily upper semi-continuous and might not attain a maximum.

In the specific case of shift spaces there does exist a simple argument to prove the variational principle which we outline now:

Let $X \subset \mathcal{A}^{\mathbb{Z}^d}$ be a shift space. Divide $\mathbb{Z}^d$ into translates of B_n and on each translate uniformly choose a pattern from $\mathcal{L}(X, B_n)$. Now translate the measure thus obtained by vectors in B_n to get a shift-invariant probability measure μ_n. A simple calculation now shows that $h_{\mu_n} \to h_{top}(X)$.

Upper semi-continuity of the entropy implies that any limit point μ of the measures μ_n (which exists because $\mathbb{P}(X)$ is compact) has the entropy greater than or equal to $h_{top}(X)$.

Further $supp(\mu) \subset X$. We are left to prove that

$$\sup_{\mu \in \mathcal{P}(X)} h_\mu \leq h_{top}(X).$$

This follows from Question 4 which implies that for any $\mu \in \mathcal{P}(X)$ we have that

$$H(\mu_{B_n}) \leq \log |\mathcal{L}(X, B_n)|.$$

A similar construction can be found in [7, Theorem 3.1] which the reader is now prepared to follow. That theorem shows another very important result that entropy of an SFT is a right-recursively enumerable number. This was proved earlier by Friedland [4] but with a slightly more complicated proof. We already know that the topological entropy of a subshift is the infimum of $\frac{1}{|B_n|}|\mathcal{L}(X, B_n)|$ however it is not possible to give an algorithm to tell whether $\mathcal{L}(X, B_n)$ is empty or not, let alone find out its cardinality. Given a forbidden list $\mathcal{F}$ one can however find the cardinality of $\text{Loc}(\mathcal{F}, B_n)$ which is the set of pattern on B_n avoiding translates of patterns from $\mathcal{F}$. These are called *locally allowed words*. It turns out that for any choice of $\mathcal{F}$ giving rise to the same shift space X we have that

$$h_{top}(X) = \inf \frac{1}{|B_n|} \log(|\text{Loc}(\mathcal{F}, B_n)|).$$

The argument is very similar to above; the only difference is that instead of taking patterns from $\mathcal{L}(X, B_n)$ we now take patterns from $\text{Loc}(\mathcal{F}, B_n)$.

Finally we have the following very important result that we will need.

Theorem 6. *If X is a nearest neighbour SFT then measures of maximal entropy of X are uniform Gibbs measures.*

This is a far-reaching generalisation of Question 5 and a simplified version of the Lanford-Ruelle theorem. A nice short proof can be found in [1, Proposition 1.19]. The argument is not difficult given the ideas above. Suppose there is a measure of maximal entropy μ of x which is not a uniform Gibbs measure. Then there must exists a finite set $B \subset \mathbb{Z}^d$ and $a \in \mathcal{L}(X, \partial B)$ such that $\mu([a]_\partial B) > 0$ and

$$\mu([\cdot]_B \ : \ [a]_{\partial B})$$

is not uniform on $c \in \mathcal{L}(X, B \cup \partial B)$ where $c|_{\partial B} = a$. Now choose n large enough such that $B \subset \{-n, \ldots, n-1, n\}^d$ and divide $\mathbb{Z}^d$ into translates of $\{-n, \ldots, n-1, n\}^d$ getting a partition $\vec{i} + \{-n, \ldots, n-1, n\}^d; \vec{i} \in I$. Take a sample x from the measure μ. For each such x, resample $x|_{\vec{i}+B}$ for $\vec{i} \in I$ if $x|_{\partial(\vec{i}+B)} = \sigma^{\vec{i}}(a)$ giving us a measure μ_n. Although μ_n is not shift-invariant you can average it out so as to make it shift-invariant. By Question 5 and prudent use of the ergodic theorem it is not hard to prove that the resulting measure has higher entropy. Thus for the full shift $\{0,1\}^{\mathbb{Z}^d}$, and $x, y \in \{0,1\}^{\mathbb{Z}^d}$ and $A \subset \mathbb{Z}^d$ a finite set we have

$$\mu([x]_A \mid [y]_{\mathbb{Z}^d \setminus A}) = \frac{1}{2^{|A|}}.$$

This is the same as choosing either 0 or 1 at each site in $\mathbb{Z}^d$ independently and with equal probability.

The converse holds in very specific situations and is easier to prove.

Question 7. *Prove that shift-invariant uniform Gibbs measures are measures of maximal entropy for the hard-core shift. Can you identify a weaker assumption that will let the proof go through?*

Hint: Notice that for the hard-core shift X given $A, B \subset \mathbb{Z}^d$ which are sufficiently separated and $x, y \in X$ there exists $z \in X$ such that $z|_A = x|_A$ and $z|_B = y|_B$.

We finish with a very interesting question by Mike Hochman. One of the great successes of (measure-theoretic) entropy was the realisation by Ornstein that (measure-theoretic) entropy was the complete invariant of isomorphism of iid processes, that is, two iid processes are isomorphic (under measure theoretic isomorphism) if and only if they have the same entropy [12]. In general though, isomorphism of processes is a wildly complicated study with no hope for such a simple criterion [3].

In a different direction, Hochman's recent results [5, 6] implied that entropy is a complete invariant for Borel isomorphism of SFTs (under a mixing assumption) modulo the periodic points. In particular the two full shift and the space of proper 3-colourings are Borel isomorphic modulo periodic points. The question now arises whether these Borel isomorphisms can be replaced by homeomorphisms.

Question 8 (Hochman—this is open though)**.** *Let X and Y be the full shift on k symbols and the space of proper $k+1$-colorings leaving out the periodic points respectively. We know that they have the same entropy from Question 1 that they have the same entropy. Is there an equivariant homeomorphism between the two?*

References

[1] R. Burton and J. E. Steif. Non-uniqueness of measures of maximal entropy for subshifts of finite type. *Ergodic Theory Dynam. Systems*, 14(2):213–235, 1994.

[2] H. Duminil-Copin, M. Gagnebin, M. Harel, I. Manolescu, and V. Tassion. The Bethe ansatz for the six-vertex and XXZ models: an exposition. *Probab. Surv.*, 15:102–130, 2018.

[3] M. D. Foreman, D. J. Rudolph, and B. Weiss. On the conjugacy relation in ergodic theory. *C. R. Math. Acad. Sci. Paris*, 343(10):653–656, 2006.

[4] S. Friedland. On the entropy of $\mathbf{Z}^d$ subshifts of finite type. *Linear Algebra Appl.*, 252:199–220, 1997.

[5] M. Hochman. Isomorphism and embedding of Borel systems on full sets. *Acta Appl. Math.*, 126:187–201, 2013.

[6] M. Hochman. Every Borel automorphism without finite invariant measures admits a two-set generator. *J. Eur. Math. Soc. (JEMS)*, 21(1):271–317, 2019.

[7] M. Hochman and T. Meyerovitch. A characterization of the entropies of multidimensional shifts of finite type. *Ann. of Math. (2)*, 171(3):2011–2038, 2010.

[8] G. Keller. *Equilibrium states in ergodic theory*, volume 42 of *London Mathematical Society Student Texts*. Cambridge University Press, Cambridge, 1998.

[9] E. H. Lieb. Residual entropy of square ice. *Physical Review*, 162(1):162, 1967.

[10] D. Lind and B. Marcus. *An introduction to symbolic dynamics and coding.* Cambridge University Press, Cambridge, 1995.

[11] M. Misiurewicz. A short proof of the variational principle for a $\mathbb{Z}_+^n$ action on a compact space. *Bull. Acad. Polon. Sci. Sér. Sci. Math. Astronom. Phys.*, 24(12):1069–1075, 1976.

[12] D. Ornstein. Bernoulli shifts with the same entropy are isomorphic. *Advances in Math.*, 4:337–352, 1970.

[13] R. Pavlov. Approximating the hard square entropy constant with probabilistic methods. *Ann. Probab.*, 40(6):2362–2399, 2012.

Einstein Institute of Mathematics, Hebrew University of Jerusalem, Israel
E-mail address: nishant.chandgotia@gmail.com