TOPOLOGICAL ENTROPY AND METRIC ENTROPY OF SETS

LIAM HODGKINSON

The majority of this talk comes from Chapter 4 of Luis Barreira's book, and from the ancient paper of Kolmogorov and Tikhimirov entitled "$\mathcal{E}$-entropy and $\mathcal{E}$-capacity of sets in functional spaces".

Topological Entropy. For any collection of sets $\mathcal{U}$ and $\mathcal{V}$, let

$$\mathcal{U} \vee \mathcal{V} = \{U \cap V : U \in \mathcal{U}, V \in \mathcal{V}\},$$

denote the *common refinement* of $\mathcal{U}$ and $\mathcal{V}$. Recall that the definition of metric entropy for a dynamical system (a transformation T that is measure-preserving with respect to a probability measure μ) is given by

$$h_\mu(T) = \sup_\xi \lim_{n\to\infty} \frac{1}{n} H_\mu\left(\bigvee_{k=0}^{n-1} T^{-k}\xi\right),$$

where the supremum is taken over all measurable partitions ξ of X. This object had a number of nice properties — in particular, it was invariant under measure-theoretic isomorphisms, and allowed for the characterisation of the complexity of many dynamical systems (making them distinguishable as well!). Recall also the concept of maximal entropy; for any measure μ and partition ξ,

$$H_\mu(\xi) \leq \log|\xi|.$$

where $|\cdot|$ here denotes the number of atoms of the partition. Aside from allowing for the formulation of 'uniform measures' as those which achieve this bound, it provides an interesting combinatorial perspective on entropy. In fact, observe that this upper bound is purely set-theoretic; it does not involve the measure μ at all. If we were to plug this upper bound into the definition of metric entropy, we arrive at a topological invariant which somewhat extends the notion to topological dynamical systems.

Definition 1. A topological dynamical system is a pair (X, T), where X is a compact Hausdorff space, and $T : X \to X$ is a continuous map. The *topological entropy* of T is given by

$$h(T) = \sup_{\mathcal{U}} \lim_{n\to\infty} \frac{1}{n} \log\left|\bigvee_{k=0}^{n-1} T^{-k}\mathcal{U}\right|,$$

where the supremum is taken over all finite open covers $\mathcal{U}$ of X, and $|\cdot|$ here denotes the cardinality of the smallest sub-cover. This limit exists because the refinements are submultiplicative in $|\cdot|$ (so $|\mathcal{U} \vee \mathcal{V}| \leq |\mathcal{U}||\mathcal{V}|$.

Like metric entropy for dynamical systems, the topological entropy enjoys the following properties:

(1) $h(T^n) = nh(T)$
(2) If T is a homeomorphism, then $h(T) = h(T^{-1})$
(3) If $T_1, T_2 : X \to X$ are continuous, and there exists a homeomorphism $\phi : X \to X$ such that $T_2 = \phi^{-1} \circ T_1 \circ \phi$, then $h(T_1) = h(T_2)$.

This last property is the preservation of topological conjugacy. In my opinion, having a topological invariant form of entropy is quite natural. In some respects, a measure of complexity of the dynamics of a continuous transformation should not require finding an invariant probability measure to obtain. For example, the logistic map does possess the invariant measure

$$\mu(A) = \int_A \frac{1}{\pi\sqrt{x(1-x)}}\, dx,$$

but the complexity of the logistic map is plain to see without drawing attention to this measure.

Furthermore, because we invoked the principle of maximal entropy to construct the topological entropy, it is to be expected that the topological entropy bounds the metric entropy over all T-invariant probability measures. But even more is true.

Theorem 2 (The Variational Principle). *If $T : X \to X$ is a continuous transformation of a compact metric space (X, d), then*

$$h(T) = \sup\{h_\mu(T) : \ \mu \text{ is a } T\text{-invariant probability measure over } X\}.$$

Proof. It should be relatively evident why $h(T) \geq h_\mu(T)$ for any T-invariant probability measure μ over X. The converse direction is immensely more difficult to show; see the $4\frac{1}{2}$ page proof in Barreira's book. □

Remark. The supremum in the variational principle will *not necessarily* be achieved by an invariant probability measure; in fact, there are many cases when it doesn't. However, if it is achieved by some measure μ, then T is ergodic with respect to μ.

While this provides a topological invariant with properties analogous to those of metric entropy, we have neither a good interpretation of this object, nor a reasonable way to calculate it in many circumstances. The objective of this note will be to follow the work of Bowen and Dinaburg in using the underlying ideas of metric entropy on sets to form a nice interpretation of topological entropy on a compact metric space (X, d). This will also provide a way of computing this object in a variety of settings, as well as show when it is likely to be finite.

Metric Entropy of Sets. The metric entropy of sets was originally conceived by Kolmogorov as a way to measure the size of certain spaces. Today, the motivation primarily stems from the desire to compare two measures on the same space. If μ and ν are two measures on X, then the total variation metric between them is given by

$$\|\mu - \nu\|_{\text{TV}} = \sup_{B \in \mathcal{B}} |\mu(B) - \nu(B)|,$$

where $\mathcal{B}$ is the Borel σ-algebra on X. The total variation metric is an example of a metric of the form

$$d_{\mathcal{F}}(\mu, \nu) = \sup_{f \in \mathcal{F}} \left| \int f d\mu - \int f d\nu \right|,$$

where $\mathcal{F}$ is a separating class of functions (here, $\mathcal{F}$ is the class of Borel-measurable indicator functions). Such metrics are ubiquitous in probability theory. Intuitively, the class $\mathcal{F}$ for the total variation metric should be quite large indeed, and this has significant repercussions. The total variation metric is rarely small, even between probability measures that should be close (for example, if μ is discrete and ν continuous, $\|\mu - \nu\|_{\text{TV}}$ will certainly never be small). Finding appropriate classes $\mathcal{F}$ with which to compare two probability measures became a field of its own in the 20th century, culminating in the Vapnik-Chervonenkis Theorem, which serves as a foundational principle of the modern theory of machine learning. To summarise, a good choice of $\mathcal{F}$ is determined by its size relative to $\mathcal{C}(X)$; the metric entropy of sets provides a general and concrete way of obtaining this.

Let (X, d) be a metric space, with a relatively compact subset A. By definition, we know that for any $\epsilon > 0$, A can be covered by finitely many balls of radius ϵ. This observation leads us to the definition of an ϵ-covering:

Definition 3. A collection of subsets $\mathcal{U}$ of A is called an *ϵ-covering* of A if the diameter of U, $d(U) \leq \epsilon$ for any $U \in \mathcal{U}$, and $X \subset \bigcup_{U \in \mathcal{U}} U$.

- Let $\mathcal{N}_\epsilon(A, d)$ denote the minimal number of sets in an ϵ-covering of X.
- The *metric entropy* of the set (A, d) is the function $\mathcal{H}_\epsilon(A, d)$ defined by

$$\mathcal{H}_\epsilon(A, d) = \log \mathcal{N}_\epsilon(A, d).$$

The metric entropy of a set is a metric invariant, in the sense that it depends only on A and d, and not on the overarching space X (this may seem obvious, but historically speaking, such an object was mindblowing at the time). In particular, the growth rate of the metric entropy of a set in ϵ provides a good way of measuring the size of a set, without requiring measures.

Remark. There are, in fact, many ways of defining the metric entropy of a set, in terms of packing numbers, internal/external covering numbers, sizes of ϵ-nets, etc. However, for any of these different definitions, the corresponding metric entropies differ by at most a constant independent of ϵ. Usually this constant is irrelevant, as we are interested primarily in the order of the metric entropy in ϵ.

Example 4. Consider the Euclidean space $(\mathbb{R}^m, \|\cdot\|_2)$, and suppose $X \subset B_r(0)$. Then

$$\mathcal{H}_\epsilon(X, \|\cdot\|_2) \leq m \log\left(\frac{2rm}{\epsilon}\right).$$

By extension, any Banach space of finite dimension will have metric entropy of order $\mathcal{O}(\log(1/\epsilon))$. The same is true for compact manifolds. Spaces with metric entropy of this order are considered *small.*

Example 5. Let $\mathcal{C}^\alpha([0,1]^m)$ denote the class of H older-continuous functions of order α on the set $[0,1]^m$. Now let $\mathcal{F}$ denote those H older-continuous functions with H older norm bounded by one. This space is a relatively compact subset of $\mathcal{C}([0,1]^m)$ (by the Arzela-Ascoli theorem). Although more challenging, its metric entropy is given by

$$\mathcal{H}_\epsilon(\mathcal{F}, \|\cdot\|_\infty) = \mathcal{O}(\epsilon^{-m/\alpha}).$$

The growth rate of this metric entropy is significantly greater than the finite-dimensional case.

Example 6. A logarithmic growth rate of the metric entropy does not imply a finite-dimensional space, however. Let Φ denote the set of entire functions in m arguments satisfying

$$|f(z_1, \ldots, z_m)| \leq C \exp\left(\sum_i \sigma_i |\Im(z_i)|^p\right)$$

for fixed $C > 0$ and σ, p. Let $\|f\|_{\infty,m} = \sup_{|z_1|,\ldots,|z_m| \leq 1} |f(z_1, \ldots, z_m)|$. Then

$$\mathcal{H}_\epsilon(\Phi, \|\cdot\|_{\infty,m}) = \mathcal{O}\left(\log\frac{1}{\epsilon}\right)^{m(p-1)/(p+1)}.$$

If $p = (m+1)/(m-1)$, then this metric entropy achieves the same growth rate as a finite-dimensional space. In essence, this function space is so regular and so small, that it can be treated *as if it was finite-dimensional.* Many other analytic function spaces possess similar growth rates as well.

The Bowen-Dinaburg Construction. Let us now use the concept of metric entropy of sets to provide a more concrete description of topological entropy for compact metric spaces. Suppose that $\mathcal{U}$ is an ϵ-covering of X. By definition, $\mathcal{N}_\epsilon(X, d) \leq |\mathcal{U}|$. Let $\mathcal{U}_n = \bigvee_{k=0}^{n-1} T^{-k}\mathcal{U}$ for each $n \geq 1$. If $x, y \in X$ lie in the same set in $\mathcal{U}_n$, then there exist sets $U_1, \ldots, U_n$ such that $T^{k-1}(x), T^{k-1}(y) \in U_k$ for each $k = 1, \ldots, n$. Therefore

$$d_{n,T}(x, y) \coloneqq \max_{k=0,\ldots,n-1} d(T^k(x), T^k(y)) < \epsilon,$$

and so $\mathcal{U}^n$ is an ϵ-covering of X under the new metric $d_{n,T}$, and $\mathcal{N}_\epsilon(X, d_{n,T}) \leq |\mathcal{U}_n|$. A similar argument provides an upper bound as well, this time involving Lebesgue numbers.

Lemma 7 (Lebesgue's Number Lemma). *For any open cover $\mathcal{U}$ of a compact metric space X, there is a number $\delta > 0$ such that every subset of X with diameter less than δ is contained in some $U \in \mathcal{U}$. Equivalently, any δ-covering of X is a refinement of an subcover $\mathcal{U}$. Any such δ is called a Lebesgue number.*

If $\mathcal{U}$ has Lebesgue number δ, then $|\mathcal{U}| \leq \mathcal{N}_\delta(X, d)$. Similarly, by the above argument, $|\mathcal{U}_n| \leq \mathcal{N}_\delta(X, d_{n,T})$. Therefore, by letting

$$\mathcal{H}_\epsilon(T) = \lim_{n\to\infty} \frac{1}{n}\mathcal{H}_\epsilon(X, d_{n,T}),$$

for any ϵ-covering $\mathcal{U}$ with Lebesgue number δ,

$$\mathcal{H}_\epsilon(T) \leq \lim_{n\to\infty} \frac{\log|\mathcal{U}_n|}{n} \leq \mathcal{H}_\delta(T),$$

which implies that

$$h(T) = \sup_{\epsilon>0} \mathcal{H}_\epsilon(T).$$

Thus, we have obtained an alternative construction of topological entropy on compact metric spaces. Concerning the finiteness of topological entropy, the supremum over ϵ may be cause for concern. Fortunately, unlike metric entropy for sets...

Lemma 8. *If T is Lipschitz and X is a compact manifold, then $\mathcal{H}_\epsilon(T)$ is bounded in ϵ.*

Proof. Since $d(T(x), T(y)) \leq Ld(x, y)$ for all $x, y \in X$ for some $L \geq 1$, it follows that

$$d(x, y) \leq L^{-n}\epsilon \quad \text{implies that} \quad d_{n,T}(x, y) \leq \epsilon.$$

Therefore

$$\mathcal{H}_\epsilon(X, d_{n,T}) \leq \mathcal{H}_{L^{-n}\epsilon}(X, d) \leq C|\log(L^{-n}\epsilon)| \leq C(n\log L + |\log\epsilon|),$$

since the metric entropy of a compact manifold is $\mathcal{O}(\log(1/\epsilon))$. It follows then that $\mathcal{H}_\epsilon(T) \leq C\log L$, which is independent of ϵ. □

Since the metric entropy for sets provides a measure of the size of a set, $\mathcal{H}_\epsilon(T)$ describes the size/complexity of the space of orbits under T. More precisely, if we can only distinguish two points that are more than ϵ distance apart, $\mathcal{H}_\epsilon(T)$ will describe the average exponential growth rate of the number of distinguishable orbits. Topological entropy is therefore a measurement of the exponential complexity of a system. I like to think of this as the rate at which trajectories are splitting apart, although the Lyapunov exponent provides a more precise description of this.

To finish up, let's calculate the topological entropy in a few example cases.

Example 9. If T is an isometry with respect to d, then $d(T^k(x), T^k(y)) = d(x, y)$, and so $\mathcal{H}_\epsilon(X, d_{n,T}) = \mathcal{H}_\epsilon(X, d)$ and $h(T) = 0$. Since any rotation is an isometry, $h(T) = 0$ in these cases (as expected).

Example 10. Consider the expanding map $E_q : \mathbb{S}^1 \to \mathbb{S}^1$ for $q > 1$ given by $E_q(x) = qx \bmod 1$. If $d(x, y) < q^{-n}$, then

$$d_{n,E_q}(x, y) = q^{n-1}d(x, y).$$

Let $\{x_0, \ldots, x_{N-1}\}$ denote a partition of $\mathbb{S}^1$ into $N = q^{n+k}$ subintervals, for some $k \geq 1$. Since $d(x_i, x_{i+1}) < q^{-n}$, it follows that $d_{n,E_q}(x_i, x_{i+1}) = q^{-k-1}$. Therefore

$$\mathcal{N}_{q^{-k-1}}(\mathbb{S}^1, d_{n,E_q}) = q^{n+k}.$$

Therefore,

$$h(E_q) = \lim_{k\to\infty} \lim_{n\to\infty} \frac{1}{n}\mathcal{H}_{q^{-k-1}}(\mathbb{S}^1, d_{n,E_q}) = \log q.$$

Now, since $h(E_q) = h_{\text{Leb}}(E_q)$, the Lebesgue measure is the E_q-invariant measure of maximal entropy.