SHIFTS AS DYNAMICAL SYSTEMS

PEYAM RYAN TABRIZIAN

Abstract. In this paper, we will discuss several important applications of symbolic dynamics in the field of dynamical systems. After recalling some theorems in analysis and dynamical systems, we will launch ourselves in the applications of symbolic dynamics to the theory of Zeta Functions, Markov Partitions, and Homoclinic Orbits. This paper will be based on chapter 6 of Lind and Marcus' book "Introduction to Symbolic Dynamics and Coding" [2]

1. Introduction

Symbolic dynamics is a powerful tool used in the study of dynamical systems, notably in the study of chaotic behavior. Its advantage lies in the fact that this technique reduces a complicated system into a set of sequences, the latter of which being much easier to analyze! We will see various instances of this simplification! This paper, in particular, is aimed to give the readers a tasteful account of the power of this method. We will briefly recall some analysis/dynamical systems facts and then talk about invariants, and finally discuss three main topics: The Zeta function, Markov partitions, and Homoclinic orbits. This paper is meant to be self-contained, although we will skip a couple of proofs!

2. Symbolics/Analysis/Dynamical Systems background

This section briefly summarizes a list of concepts needed before we can tackle the concept of symbolic dynamics.

2.1. A couple of definitions from Symbolic Theory.

First, we need a couple of definitions:

Definition 1. *An **alphabet** $\mathcal{A}$ is a set of symbols*

The most common alphabet is $[10] := \{0, 1, 2, \cdots, 8, 9\}$, and in general, $[n] := \{0, 1, 2, \cdots, n-1\}$

Definition 2. *Given an alphabet $\mathcal{A}$, the **full shift space** is $M = \mathcal{A}^{\mathbb{Z}}$ (i.e. the space of sequences from $\mathbb{Z}$ into $\mathcal{A}$)*

In particular, $\mathbf{X}_{[n]} = [n]^{\mathbb{Z}}$.

Definition 3. *A **block** over $\mathcal{A}$ is a finite sequence of symbols from $\mathcal{A}$. An **n-block** is a block of length n*

Definition 4. *If X is a subset of the full shift space, let $\mathcal{B}_n(X)$ be the set of all n-blocks that occur in points in X. Then **the language of X** is the union of all $\mathcal{B}_n(X)$, where $n \in \mathbb{N}$. We say that a block u is **allowed** in X when $u \in \mathcal{B}(X)$*

Date: Thursday, December 2nd, 2010.

Definition 5. *If $\mathcal{F}$ is a collection of blocks over $\mathcal{A}$, then $X_{\mathcal{F}}$ is the subset of $\mathcal{A}^{\mathbb{Z}}$ which do not contain any block in $\mathcal{F}$. Elements in $\mathcal{F}$ are called* ***forbidden blocks***

Definition 6. *A* ***shift space*** *is a subset X of $\mathcal{A}^{\mathbb{Z}}$ such that $X = X_{\mathcal{F}}$ for some collection $\mathcal{F}$ of forbidden blocks over $\mathcal{A}$*

Definition 7. *Given a sequence $\boldsymbol{x} \in \mathcal{A}^{\mathbb{Z}}$, we define the* ***shift map*** *to be the map $\sigma : \mathcal{A}^{\mathbb{Z}} \longrightarrow \mathcal{A}^{\mathbb{Z}}$ defined by: $(\sigma(\boldsymbol{x}))_n = x_{n+1}$ (in other words, σ shifts a sequence to the left)*

Definition 8. *If X is a shift space, then $\sigma_{|X}$ is the restriction of σ to X*

Definition 9. *If $\sigma = \sigma_{\big|X_{[n]}}$, for some n then σ is called a* ***shift of finite type***

Definition 10. *Given a sequence $\boldsymbol{x} \in \mathcal{A}^{\mathbb{Z}}$, and given $k \in \mathbb{N}$, we define $x_{[-k,k]}$ to be the finite word (x_i), with $-k \leq i \leq k$. It is called the* ***central k-block***

2.2. **List of analysis concepts that should be known to the readers.** The following concepts should already be known to the readers, so we won't recall them. Instead, we will point out how those concepts are used in symbolic dynamics.

(1) Metric spaces
(2) Normed spaces
(3) $\mathbb{T}^2 \cong [0,1)^2$
(4) Convergence in metric spaces
(5) Continuity and homeomorphisms
(6) Compactness
(7) Open/Closed sets, and Bases of topological spaces
(8) Bolzano-Weierstrass Theorem
(9) Heine-Borel Theorem
(10) Baire Category Theorem

2.3. **Applications of the above concepts to symbolic dynamics.** For symbolic dynamics, the most important example of a metric space is the **shift space**.

Example 1 (Shift Spaces)**.** *Let M be a shift space, and put:*

$$\rho(\boldsymbol{x}, \boldsymbol{y}) = \begin{cases} 2^{-k} & \text{if } k \text{ is maximal such that } x_{[-k,k]} = y_{[-k,k]} \\ 0 & \text{if } \boldsymbol{x} = \boldsymbol{y} \end{cases}$$

In other words, $\boldsymbol{x}$ and $\boldsymbol{y}$ are 2^{-k}-close if their central $(2k+1)$-blocks agree. The proof that ρ is a metric is fairly trivial, and will be omitted. The point is that (M, ρ) becomes a metric space, and we can apply the theory of metric spaces to study shift-spaces. The metric captures the idea that two points are close iff they agree on a large enough central block.

Remark 1. *We can use other metrics on M such as:*

$$\rho(\boldsymbol{x}, \boldsymbol{y}) = \sum_{k=-\infty}^{k=\infty} \frac{|x_k - y_k|}{2^{|k|}}$$

or

$$\rho(\boldsymbol{x}, \boldsymbol{y}) = \frac{1}{k+2}, \text{ where } k \text{ is maximal such that } x_{[-k,k]} = y_{[-k,k]}$$

But they are all equivalent!

Example 2 (Convergence in shift spaces)**.** *From our metric above, we see that a sequence of points in a shift space converges exactly when, for each $k \geq 0$, the central $(2k+1)-$blocks spabilize starting at some element of the sequence. More precisely, if X is a shift space, and $x^{(n)}$ be a sequence in X (so $x^{(n)}$ is a sequence for* ***each*** *n). Then we have that $x^{(n)} \to x \in X$ when, for each $k \geq 0$, there is an n_k such that $x^{(n)}_{[-k,k]} = x_{[-k,k]}$ for all $n \geq n_k$. For example, if $x^{(n)}$ is the sequence where the block 10^n is repeated infinitely many times (i.e. $x^{(1)} = ...101010.101010...$, $x^{(2)} = ...100100100.100100100...$), then $x^{(n)} \to y$, where $y = 0.100...$*

Example 3 (Continuity of $\sigma_{|_X}$)**.** *Let X be a shift space If two points in X is close, they agree on a large central block, hence their images under $\sigma_{|_X}$ also agree on a large central block. Thus $\sigma_{|_X}$ is continuous.*

Example 4 (Compactness of shift spaces)**.** *Let $M = X$, a shift space. We claim that M is compact. To see this, given a sequence $x^{(n)}$ in X, construct a convergent subsequence using the Cantor diagonalization argument. More precisely, for $k \geq 1$, inductively find a decreasing sequence S_k of infinite subsets of positive integers so that all blocks $x^{(n)}_{[-k,k]}$ are equal for all $n \in S_k$. Define x to be the point with $x_{[-k,k]} = x_{[-k,k]}$ for all $n \in S_k$, and inductively define n_k as the smallest element of S_k which exceeds n_{k-1}. Then $x \in X$, and $x^{(n_k)}$ converges to x as $k \to \infty$ by construction!*

Example 5 (Basis for shift spaces)**.** *If X is a shift space, one good basis for X is what we call the set of* ***cylinders****. More precisely, if $u \in \mathcal{B}(X)$ (the set of blocks in X) and $k \in \mathbb{N}$, define the* ***cylinder set*** *$C_k(u)$ as:*

$$C_k(u) = C_k^X(u) = \left\{x \in X \mid x_{[k,k+|u|-1]=u}\right\}$$

i.e. $C_k(u)$ is the set of points in which the block u occurs starting at position k. Then it can be shown that the set $\mathcal{C} = \{C_k(u) \mid k \in \mathbb{N}, u \in \mathbb{B}(X)\}$ forms a basis of X.

Finally, the next result characterizes shift spaces:

Theorem. *A subset of $\mathcal{A}^{\mathbb{Z}}$ is a shift space iff it is invariant under σ and compact*

Proof. ($\Longrightarrow$) It is easy to check that shift spaces are shift-invariant and compact
($\Longleftarrow$) Suppose that $X \subseteq \mathcal{A}^{\mathbb{Z}}$ is shift-invariant and compact. Then X is closed, and $\mathcal{A}^{\mathbb{Z}} \backslash X$ is open. Hence, for each $y \in \mathcal{A}^{\mathbb{Z}} \backslash X$, there is a $k = k(y)$ such that if $u_y = y_{[-k,k]}$, then $C_{-k}^{\mathcal{A}^{\mathbb{Z}}}(u_y) \subseteq \mathcal{A}^{\mathbb{Z}} \backslash X$. Now let $\mathcal{F} = \left\{u_y \mid y \in \mathcal{A}^{\mathbb{Z}} \backslash X\right\}$ Then it is easy to verify using the shift-invariance of X that $X = \mathbf{X}_{\mathcal{F}}$, so X is a shift space $\square$

2.4. **Dynamical Systems Background.** This is a very brief subsection on dynamical systems. We will mainly show how we can relate dynamical systems and symbolic theory. Briefly said, the subject of dynamical systems studies how a given system behaves throughout time, but studying discrete or continuous iterates.

Definition 11. *A* ***dynamical system*** *(M, ϕ) consists of a compact metric space M together with a continous map $\phi : M \to M$, which is usually a homeomorphism.*

Prominent examples include the **irrational rotation on the torus** and **the doubling map**.

To relate this subject with symbolic theory, we need the definition of a **sliding block code**

Definition 12. *Let X be a shift space over $\mathcal{A}$, and define $\Phi : \mathcal{B}_{m+n+1}(\mathcal{A}) \to \aleph$ be a block map. Then the map $\phi : X \to \aleph$ defined by: $\phi(x) = y$, where $y_i = \Phi(x_{i-m}x_{i-m+1}\cdots x_{i+n}) = \Phi(x_{[i-m,i+n]})$ is called the **sliding block code with memory m and anticipation n induced by** Φ. We will denote the formation of ϕ from Φ by $\phi = \Phi_\infty^{[-m,n]}$, or more simply by $\phi = \Phi_\infty$.*

Basically, what we're doing is *identifying* blocks!

The importance of sliding block codes stems from the following very important fact:

Proposition. *Let X and Y be shift spaces. If $\phi : X \to Y$ is a sliding block code, then $\phi \circ \sigma_X = \sigma_Y \circ \phi$, i.e. σ_X and σ_Y are conjugate*

One can show, in the same way as we did for σ, that $\phi = \Phi_\infty$ is continuous. In particular, we get the following important example:

Example 6. *Let $M = X$, a shift space, and $\phi = \Phi_\infty : X \to X$ a sliding block code. Then (M, ϕ) is a dynamical system.*

Given those definitions and examples of dynamical systems, one is led to ask the following questions:

(1) For each point $x \in M$, are there periodic points arbitrarily close to x?
(2) How are the points of an orbit distributed throughout M?
(3) Given a subset U of M, how do the sets $\phi^n(U)$ spread throughout M, as $n \to \infty$?
(4) When are two dynamical systems **conjugate**?

Definition 13. *A **homomorphism** θ from (M, ϕ) to (N, ψ) is a continuous function $\theta : M \to N$ such that $\psi \circ \theta = \theta \circ \phi$*

Definition 14. *Two dynamical systems (M, ϕ) and (N, ψ) are **topological conjugate** if there exists a homomorphism θ between them that is also a hom**e**omorphism*

Note that, because we defined dynamical systems to be defined on compact spaces, the fact that the inverse is continuous is already given to us!

The last problem is very important. Basically, topological conjugate systems have the same behavior. This is a crucial property, because usually one dynamical system is hard to analyze, but if we can show that that system is conjugate to an easier-to-study system, then we can deduce all the needed properties for the hard system from the easy system! This is even more apparent in the study of homoclinic orbits (see last section).

The following important result characterizes the homomorphisms from one shift dynamical system to another one:

Theorem (Curtis-Lyndon-Hedlund Theorem)**.** *Suppose that (X, σ_X) and (Y, σ_Y) are shift dynamical systems, and that $\theta : X \to Y$ is a (not necessarily continuous) function. Then θ is a sliding block code iff it is a homomorphism.*

Let's prove this theorem because it is fairly short and is a good illustration of 'cylinder sets' in action.

Proof. ($\Leftarrow$) One can show that sliding block codes are continuous in the same way we showed that σ_X is continuous. And clearly sliding block codes are shift commuting, so any sliding block code is a homomorphism.
($\Rightarrow$) Conversely, suppose that θ is a homomorphism. Let $\mathcal{A}$ be the alphabet of X and $\aleph$ be the alphabet of Y. For each $b \in \aleph$, let:

$$C_0(b) = \{y \in Y \mid y_0 = b\}$$

The sets $C_0(b)$ are disjoint and compact, so their inverse images $E_b = \theta^{-1}(C_0(b))$ are also disjoint and compact in X. Hence, there is a $\delta > 0$ such that points in different sets E_b are at least a distance of δ apart. Now choose n with $2^{-n} < \delta$. Then any pair of points $x, x' \in X$ such that $x_{[-n,n]} = x'_{[-n,n]}$ must lie in the same set E_b, so that $\theta(x)_0 = b = \theta(x')_0$. Thus, the 0th coordinate of $\theta(x)$ only depends on the central $(2n+1)-$block of x. Now, simply define the $(2N+1)-$block map Φ by $\Phi(w) = \theta(x)_0$, where x is any point in X. Then it is straightforward to check that $\theta = \Phi_\infty^{[-N,N]}$, so θ is a sliding block code. $\square$

3. Invariants

Showing that two dynamical systems are conjugate is fairly easy (in theory): just find an appropriate conjugacy! What is **much** harder to show is that two dynamical systems are **NOT** conjugate. To show the latter, we usually use an object that is called an *invariant*. More precisely, it's a function f from the set of dynamical systems to another space, usually $\mathbb{R}$, such that, when ϕ and ψ are conjugate, then $f(\phi) = f(\psi)$.
Examples of such invariants include:

(1) p_n, the number of periodic points of period n
(2) q_n, the number of periodic points of *least* period n
(3) Topological transitivity (here, the range space is { yes, no })
(4) Topological mixing (ditto)
(5) Topological entropy

We will not define those concepts, as they should already be known to the reader. Instead, we will prove a much more interesting result, for which we will need a preliminary definition:

Definition 15. *A shift space X is called **irreducible** if, for every ordered pair of blocks $u, v \in \mathcal{B}(X)$, there is a $w \in \mathcal{B}(X)$ with $uwv \in \mathcal{B}(X)$*

Basically, this says that, given two blocks, there is another block joining them. Notice that this is very similar to the concept of topological transitivity, which says that for every pair of open sets U and V, we can go from U to V with ϕ. In fact, we have the following result, which nicely illustrates everything that we discussed in the previous section:

Proposition. *For a shift dynamical system (X, σ_X), topological transitivity of σ_X is equivalent as irreducibility of X*

This really says that all the information that σ tells us is basically encoded in X!

Proof. ($\Leftarrow$) Suppose X is irreducible, and let U and V be a pair of nonempty open sets in X. We must show that there is an $n > 0$ such that $\sigma_X^n(U) \cap V \neq \emptyset$. Since

the cylinder sets form a basis of X, we will first consider a pair $C_0(u)$, $C_0(v)$ for blocks u, v. Since X is irreducible, there is a block w such that $uwv \in \mathcal{B}(X)$. Hence, if $n = |uw|$ (the size of the string uw), then $C_0(u) \cap \sigma_X^{-n}(C_0(v)) \neq \emptyset$, that is, $\sigma_X^n(C_0(u)) \cap (C_0(v)) \neq \emptyset$, so we get topological transitivity for cylinder sets. But, then, we get topological transitivity of σ_X, essentially because the cylinder sets form a basis of X!
($\Rightarrow$) Suppose that σ_X is topologically transitive. Then, given an ordered pair u, v of allowed blocks, it follows that there is an $n > |u|$ with $\sigma_X^n(C_0(u)) \cap (C_0(v)) \neq \emptyset$, that is, $C_0(u) \cap \sigma_X^{-n}(C_0(v)) \neq \emptyset$. Then, if z is in the latter set, we get $z_{[0,n+|v|-1]} = uz_{[|u|,n-1]}v = uwv$ is allowed in X, proving that X is irreducible. $\square$

Finally, similar to the proof above, it turns out that if X is a shift space with shift map σ_X, then the topological entropy of $h(\sigma_X)$ equals the entropy $h(X)$ of the shift space X. For this reason, proving that a dynamical system (M, ϕ) is conjugate to a shift is usually very powerful. It is much easier to calculate $h(X)$ than to calculate the entropy of ϕ.

4. The Zeta Function

One of the most powerful invariants in the **Zeta function**. It is to dynamical systems what the Jordan Canonical Form is to linear algebra. More precisely, two dynamical systems are conjugate if and only if they have the same Zeta function. Unfortunately, this function is very hard to compute!

Definition 16. *Let (M, ϕ) be a dynamical system with $p_n(\phi) < \infty$ for all $n \geq 1$. The **Zeta function** $\zeta_\phi(t)$ is defined as:*

$$\zeta_\phi(t) = \exp\left(\sum_{n=1}^{\infty} \frac{p_n(\phi)}{n} t^n\right)$$

Once we know $p_n(t)$, we can easily calculate $\zeta_\phi(t)$. But notice that the converse holds too, because:

$$\ln\left(\zeta_\phi(t)\right) = \sum_{n=1}^{\infty} \frac{p_n(\phi)}{n} t^n$$

And hence, by Taylor's formula, we get:

$$\frac{d^n}{dt^n} \ln \zeta_\phi(t) \mid_{t=0} = n! \frac{p_n(\phi)}{n} = (n-1)! p_n(\phi)$$

So studying $\zeta_\phi(t)$ gives us a lot of valuable information about periodic points of the system!

Let's see some examples of how to calculate zeta functions!

Example 7. *Let $M = X_{[2]}$ and $\phi = \sigma_{[2]}$ (this means that $\mathcal{A}$ has only two elements, 0 and 1). Then one can easily show that $p_n(\phi) = 2^n$ and so we get:*

$$\begin{aligned}\zeta_\phi(t) &= \exp\left(\sum_{n=1}^{\infty} \frac{2^n}{n} t^n\right)\\ &= \exp\left(\sum_{n=1}^{\infty} \frac{(2t)^n}{n}\right)\\ &= \exp\left(-\ln(1-2t)\right)\\ &= \frac{1}{1-2t}\end{aligned}$$

So $\boxed{\zeta_\phi(t) = \frac{1}{1-2t}}$

Example 8. *Let* $A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$, $M = X_A$, *and* $\phi = \sigma_A$ *(basically,* ϕ *is multiplication by* A*). Then, one can show that* $p_n(\phi) = trA^n = \lambda^n + \mu^n$, *with* $\lambda = \frac{1+\sqrt{5}}{2}$ *and* $\mu = \frac{1-\sqrt{5}}{2}$ *being the eigenvalues of* A*. Then, we get:*

$$\begin{aligned}\zeta_\phi(t) &= \exp\left(\sum_{n=1}^{\infty} \frac{\lambda^n + \mu^n}{n} t^n\right)\\ &= \exp\left(\sum_{n=1}^{\infty} \frac{(\lambda t)^n}{n} + \sum_{n=1}^{\infty} \frac{(\mu t)^n}{n}\right)\\ &= \exp\left(-\ln(1-\lambda t) - \ln(1-\mu t)\right)\\ &= \frac{1}{(1-\lambda t)(1-\mu t)}\\ &= \frac{1}{1-t-t^2}\end{aligned}$$

You may have noticed that the zeta function in the above two examples is a reciprocal of a polynomial! In fact, this is always true for shifts of finite type (i.e. shifts on a finite alphabet):

Theorem. *Let* A *be an* $r \times r$ *nonnegative integer matrix, and* $\chi_A(t)$ *its characteristic polynomial, and* σ_A *the associated shift map. Then:*

$$\zeta_{\sigma_A}(t) = \frac{1}{t^r \chi_A(t^{-1})} = \frac{1}{det(I - tA)}$$

Thus, the zeta function of a shift of finite type is the reciprocal of a polynomial!

Proof. The proof is very cute, and similar to the computation in the above example. More precisely, let the roots of $\chi_A(t)$ be $\lambda_1, \cdots, \lambda_r$, listed with multiplicity. Then, we have that $p_n(\sigma_A) = trA^n = \lambda_1^n + \cdots + \lambda_r^n$, and so:

$$\begin{aligned}\zeta_{\sigma_A}(t) &= \exp\left(\sum_{n=1}^{\infty} \frac{\lambda_1^n + \cdots + \lambda_r^n}{n} t^n\right)\\ &= \exp\left(\sum_{n=1}^{\infty} \frac{(\lambda_1 t)^n}{n} + \cdots \sum_{n=1}^{\infty} \frac{(\lambda_r t)^n}{n}\right)\\ &= \frac{1}{1-\lambda_1 t} \times \cdots \times \frac{1}{1-\lambda_r t}\end{aligned}$$

Now, since:

$$\chi_A(u) = (u-\lambda_1)\cdots(u-\lambda_r) = u^r(1-\lambda_1 u^{-1})\cdots(1-\lambda_r u^{-1}) = u^r det(I-u^{-1}A)$$

substituting t^{-1} for u shows that:

$$\chi_A(t^{-1}) = t^{-r}(1-\lambda_1 t)\cdots(1-\lambda_r t) = u^r det(I-tA)$$

And hence:

$$\zeta_{\sigma_A}(t) = \frac{1}{t^r \chi_A(t^{-1})} = \frac{1}{det(I-tA)}$$

□

A very easy, but important corollary is the following. It tells us exactly how much information is encoded in the Riemann Zeta function:

Corollary 1. *Let A be a nonegative integer matrix. Then each of the following determines the other three:*

(1) $\zeta_{\sigma_A}(t)$
(2) $sp^{\times}(A)$ *(the set of nonzero eigenvalues of A)*
(3) $\{p_n(\sigma_A)\}_{n\geq 1}$
(4) $\{q_n(\sigma_A)\}_{n\geq 1}$

5. Markov Partitions

5.1. **Introduction and an example.** One of the main uses of symbolic dynamics is representing other dynamical systems. This is useful because some dynamical systems are very hard to analyze, but their representations are much easier to analyze, and moreover give valuable information about the original system.

Let's begin with the idea of symbolic representations. Suppose we want to study a dynamical system (M, ϕ). To simplify things further, suppose that ϕ **is invertible**, so that we can use both positive and negative iterates. If $y \in M$, we can describe the orbit $\{\phi^n(y)\}_{n\in\mathbb{Z}}$ in the following way: Divide M into a finite number of pieces $E_0, E_1, \cdots, E_{r-1}$, and then track the orbit of y by keeping a record on which of thesepieces $\phi^n(y)$ lands in. This yields a corresponding sequence $\mathbf{x} = \cdots x_{-1}x_0x_1\cdots \in \{0,1,\cdots,r-1\}^{\mathbb{Z}} := \mathbf{X}_{[r]}$ defined by $\phi^n(y) \in E_{x_n}$ for $n \in \mathbb{Z}$. We're basically studying the behavior of x under iteration of ϕ
So for every $y \in M$, we get a symbolic point x, and by definition, $\phi(y)$ corresponds to $\sigma(x)$. If we're lucky (and in most cases we are), this correspondence is one-to-one, which is amazing because then the study of (M, ϕ) is just reduced to the study

of sequences!
For non-invertible dynamical systems, the best we can hope for is getting one-sided sequences (where n is in $\mathbb{N}$, not in $\mathbb{Z}$), but this doesn't change the previous comment!

Example 9. *Let $M = \mathbb{T}$, and $\phi : M \to M$ dfined by $\phi(y) = 10y$ (mod 1). Subdivide M into 10 equal subintervals $E_j = [\frac{j}{10}, \frac{j+1}{10})$ for $j = 0, 1, \cdots, 9$, so the alphabet here is $\mathcal{A} = \{0, 1, \cdots, 9\}$. In this case, y corresponds precisely to the sequence of digits in its decimal expansion (considering $M \cong [0, 1)$). The action of ϕ corresponds to the (onse-sided) shift map, i.e. multiplying a real number by 10 shifts the decimal digits to the left by one and deletes the left-most digit*

5.2. **More detailed study and Markov Partitions.** Let's study symbolic representations in more detail! This will lead to the idea of a Markov partition.

Definition 17. *A **topological partition** of a metric space M is a finite collection $\mathcal{P} = \{P_0, P_1, \cdots, P_{r-1}\}$ of disjoint open sets whose closures $\bar{P}_j$ cover M*

Definition 18. *Let (M, ϕ) a (not necessarily invertible) dynamical system and $\mathcal{P}$ a topological partition of M. Let $\mathcal{A} = \{0, 1, \cdots, r-1\}$. Then a word $w = a_1 a_2 \cdots a_n$ is **allowed for** $\mathcal{P}, \phi$ if $\bigcap_{j=1}^{n} \phi^{-1}\left(P_{a_j}\right) \neq \emptyset$*

Definition 19. *Let $\mathcal{L}_{\mathcal{P},\phi}$ be the collection of all allowed words for $\mathcal{P}, \phi$.*

One can check that $\mathcal{L}_{\mathcal{P},\phi}$ is the language of a shift space, hence there is a unique shift space $\mathbf{X}_{\mathcal{P},\phi}$ whose language is $\mathcal{L}_{\mathcal{P},\phi}$.

Definition 20. *$\boldsymbol{X}_{\mathcal{P},\phi}$ is called the **symbolic dynamical system corresponding to** $\mathcal{P}, \phi$. If ϕ is not necessarily invertible, then the one-sided shift space $\boldsymbol{X}^{+}_{\mathcal{P},\phi}$ is the **one-sided symbolic dynamical system corresponding to** $\mathcal{P}, \phi$.*

Example 10. *Again, let $M = \mathbb{T}$ and ϕ be as in the previous example. Let $\mathcal{P} = \left\{\left(0, \frac{1}{10}\right), \left(\frac{1}{10}, \frac{2}{10}\right), \cdots, \left(\frac{9}{10}, 1\right)\right\}$, and $\mathcal{A} = \{0, 1, \cdots, 9\}$. Then $\mathcal{L}_{\mathcal{P},\phi}$ is the set of all words over $\mathcal{A}$, so $\boldsymbol{X}^{+}_{\mathcal{P},\phi}$ is the full one-sided 10-shift $\boldsymbol{X}^{+}_{[10]}$*

Example 11. *Again, let $M = \mathbb{T}$ and ϕ be the identity map. Let $\mathcal{P}$, and $\mathcal{A}$ be as above. Then Then $\mathcal{L}_{\mathcal{P},\phi}$ only contains words of the form a^n wher $a \in \mathcal{A}$ and $n \geq 1$, so $\boldsymbol{X}_{\mathcal{P},\phi}$ just contains 10 points labeled $0^\infty, 1^\infty, \cdots, 9^\infty$.*

As the above example shows, the above method 'fails' for some examples, i.e. it does not give any information about the underlying dynamical system. Now we'll see a general case when the above method works! Again, suppose that we have an *invertible* dynamical system (M, ϕ). Let $\mathcal{P} = \{P_0, P_1, \cdots, P_{n-1}\}$ be a topological partition of M. For each $x \in \mathbf{X}_{\mathcal{P},\phi}$ and $n \geq 0$, consider the corresponding nonempty open set:

$$D_n(x) = \bigcap_{k=-n}^{n} \phi^{-k}\left(P_{x_k}\right) \subseteq M$$

The closures $\bar{D}_n(x)$ of these sets are compact and decrease with n, so that $\bar{D}_0(x) \supseteq \bar{D}_1(x) \supseteq \bar{D}_2(x) \supseteq \cdots$. Hence, $\bigcap_{k=-n}^{n} \bar{D}_n(x) \neq \emptyset$. Now, ideally we'd like to have a one-to-one correspondence between points in $\mathbf{X}_{\mathcal{P},\phi}$ and points in M, so the above intersection should contain only one point! This leads to the following definition of a Markov partition:

Definition 21. *Let (M, ϕ) be an invertible dynamical system. A topological partition $\mathcal{P} = \{P_0, P_1, \cdots, P_{r-1}\}$ of M gives a* ***symbolic representation*** *of (M, ϕ) if for every $x \in \boldsymbol{X}_{\mathcal{P},\phi}$ the intersection $\bigcap_{n=0}^{\infty} \bar{D}_n(x)$ consists of exactly one point. We call $\mathcal{P}$ a* ***Markov partition*** *for (M, ϕ) if $\mathcal{P}$ gives a symbolic representation of (M, ϕ) and furthermore $\boldsymbol{X}_{\mathcal{P},\phi}$ is a shift of finite type.*

There is an analog of this for not necessarily invertible dynamical systems, if we replace $D_n(x)$ with $D_n^+(x) = \bigcap_{n=0}^{\infty} \phi^{-k}\left(P_{x_k}\right)(x)$ and $\mathbf{X}_{\mathcal{P},\phi}$ with $\mathbf{X}_{\mathcal{P},\phi}^+$. This gives us the definition of a **one-sided Markov partition**.

Example 12. *The topological partition $\mathcal{P} = \left\{\left(0, \frac{1}{10}\right), \cdots, \left(\frac{9}{10}, 1\right)\right\}$ in the above example is a one-sided Markov partition for that dynamical system.*

Suppose that $\mathcal{P}$ gives a symbolic representation of the invertible dynamical system (M, ϕ). Then there is a natural map $\phi : X = \mathbf{X}_{\mathcal{P},\phi}$ which maps $x = \cdots x_{-1}x_0x_1 \cdots$ to the unique point $\pi(x)$ in $\bigcap_{n=0}^{\infty} \bar{D}_n(x)$.

Definition 22. *We call x a* ***symbolic representation*** *of $\pi(x)$*

Using that $D_{n+1}(\sigma x) \subseteq \phi\left(D_n(x)\right) \subseteq D_{n-1}(\sigma x)$, we have the following commuting diagram, which is basically why we even bothered to define π:

$$\begin{array}{ccc} \mathbf{X}_{\mathcal{P},\phi} & \xrightarrow{\sigma} & \mathbf{X}_{\mathcal{P},\phi} \\ \pi \downarrow & & \pi \downarrow \\ M & \xrightarrow{\phi} & M \end{array}$$

Basically, we identify ϕ with σ via π.

The following result shows that π is also continuous and onto and therefore a factor map from $(\mathbf{X}_{\mathcal{P},\phi}, \sigma)$ to (M, ϕ). We will skip the proof, simply noting that it uses the fact that diam $\left(\bar{D}_n\right)$ shrinks to 0 to show that π is continuous, and the Baire Category Theorem to show that π is onto!

Proposition. *Let $\mathcal{P}$ give a symbolic representation of the invertible dynamical system (M, ϕ), and let $\pi : \boldsymbol{X}_{\mathcal{P},\phi} \to M$ be the map defined above. Then π is a factor map from $(\boldsymbol{X}_{\mathcal{P},\phi}, \sigma)$ to (M, ϕ)*

Note that there is a one-sided/non-invertible version as well!

Notice that this commutative diagram shows that, whenever we have a Markov partition for a dynamical system, we can use symbolic dynamics to study that system. The next proposition summarizes this result:

Proposition. *Let $\mathcal{P}$ give a symbolic representation of the invertible dynamical system (M, ϕ). For each of the following, if $(\boldsymbol{X}_{\mathcal{P},\phi}, \sigma)$ has the property, then so does (M, ϕ).*

(1) Topological transitivity
(2) Topological mixing
(3) The set of periodic orbits is dense

And Markov partitions **do** exist in nature! Adler and Weiss [1] discovered that certain mappings from the torus to itself, called **hyperbolic toral automorphisms** have Markov partitions, and in fact these partitions are parallelograms. One famous example is the **Arnold Cat Map**:

Example 13 (Arnold Cat Map)**.** *Let $M = \mathbb{T}^2$ and $\phi((x,y)) = A(x,y)$, where addition is addition modulo 1 and:*

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

Notice that the following diagram commutes:

$$\begin{array}{ccc} \mathbb{R}^2 & \xrightarrow{A} & \mathbb{R}^2 \\ \pi \downarrow & & \pi \downarrow \\ \mathbb{T}^2 & \xrightarrow{\phi} & \mathbb{T}^2 \end{array}$$

where $\pi : \mathbb{R}^2 \to \mathbb{T}^2$ is the natural map. Hence, in this example, we will usually 'identify' ϕ and A.

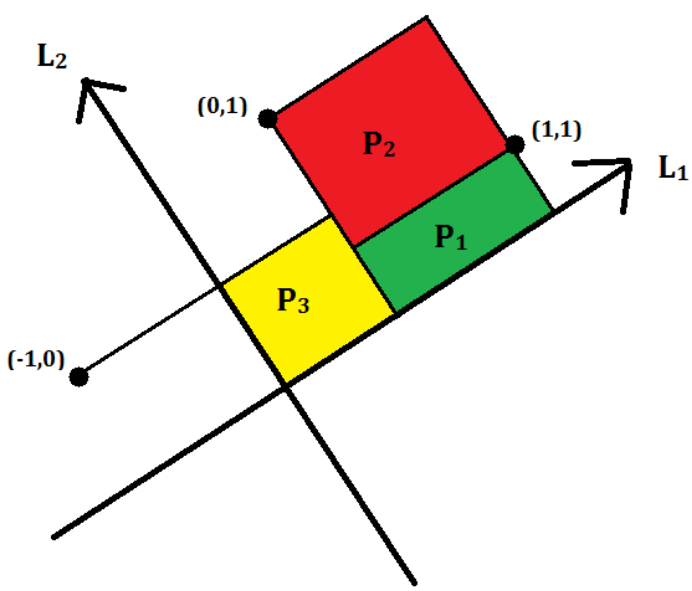
Notice that A has two eigenvalues $\lambda_+ = \frac{1+\sqrt{5}}{2}$ and $\lambda_- = \frac{1-\sqrt{5}}{2}$. Notice that $|\lambda_+| > 1$ and $|\lambda_-| < 1$, and for this reason this dynamical system is called ***hyperbolic****. Moreover, we let:*

$$L_1 = E_{\lambda_+} = span\left\{\begin{bmatrix} 1 \\ -\frac{1-\sqrt{5}}{2} \end{bmatrix}\right\}$$

$$L_2 = E_{\lambda_-} = span\left\{\begin{bmatrix} 1 \\ -\frac{1+\sqrt{5}}{2} \end{bmatrix}\right\}$$

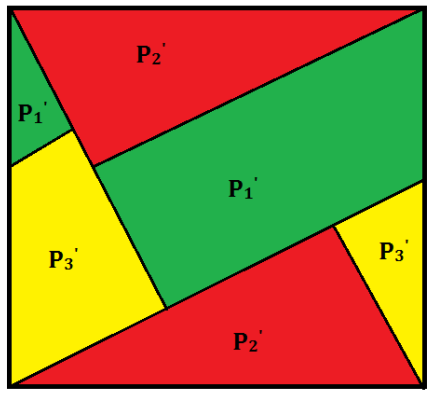
Then, a Markov partition of the system is given as follows: Let P_1, P_2, P_3 be rectangles in $\mathbb{R}^2$ as in the following picture:

as Dynamical Systems/Pre-Markov.png



And let $P_i' = \pi(P_i)$ for $i = 1, 2, 3$. One can check that $\{P_1', P_2', P_3'\}$ forms a Markov partition for ϕ, and we get the following picture:

as Dynamical Systems/Markov Partition.png



6. Homoclinic Orbits

Finally, one of the nicest applications of symbolic dynamics is their use in the study of homoclinic orbits. In this section, we will define homoclinic orbits and **briefly** discuss how symbolic dynamics are used in this problem. For more information, consult my other paper "Homoclinic Orbits and Chaotic Behavior in Classical Mechanics" [3].

The setting is as follows:

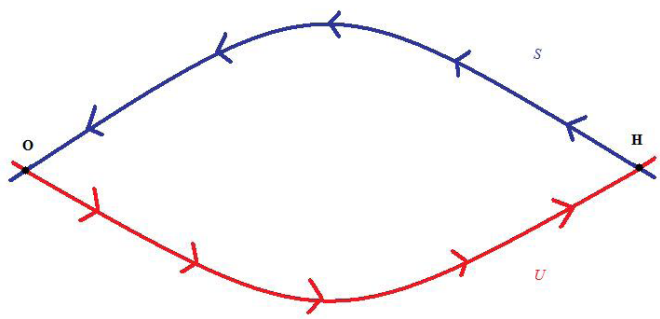
Let ϕ be an area-preserving C^1-diffeomorphism and O a hyperbolic fixed point of ϕ. Area-preserving means that $D\phi$ has detrminant 1 everywhere. Hyperbolic here means that $D\phi(O)$ has real eigenvalues λ_1 and λ_2 with $0 < |\lambda_1| < 1 < |\lambda_2|$. This guarantees the existence of a stable manifold S and an unstable manifold U for ϕ.

Now assume that U and S further intersect in a point H, called **homoclinic point**. More precisely, we have:

Definition 23. *If $H \neq p$ and $\phi^k(H)$ approaches one periodic orbit $\{p, \phi(p), \phi^2(p), \ldots \phi^m(p) = p\}$ as $k \to \pm\infty$, then H is called a **homoclinic point***

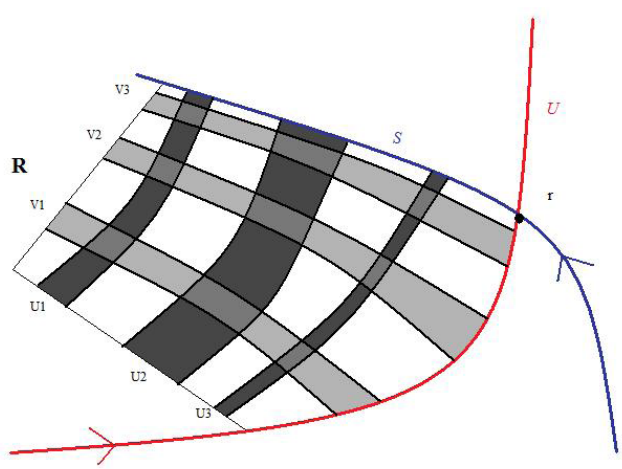
So the situation is more or less as in the following picture:

as Dynamical Systems/Homoclinic Point.png



For the rest of this, I won't go through the details (all of which can be found in my paper). The main idea that we want to study the behavior of ϕ near H, and for this we construct a quadrilateral R near H and consider the set of points $D(\tilde{\phi})$ which never escape R, and define $\tilde{\phi} = \phi_{|_R}$. Those points are precisely the intersections of some horizontal and vertical strips U_k resp. V_k, as in the following picture:

as Dynamical Systems/Strips.png



Now here comes the main theorem in this section, which is where symbolic dynamics come into play:

Theorem. *There exists an invariant set I in R a number $N \in \mathbb{N}$ and a homeomorphism $\tau : \mathcal{A}^{\mathbb{Z}} \to I$, where $\mathcal{A} = \{0, 1, 2, \cdots, N-1\}$ such that:*

$$\tilde{\phi} \circ \tau = \tau \circ \sigma$$

where σ is a shift of finite type.

And from this theorem, we get the following amazing consequences:

Corollary 2. *Homoclinic points are dense in R*

Corollary 3. *Periodic orbits are dense in R*

Corollary 4. *Points in R enjoy the property of* ***topological entropy*** *, that is, in any neighborhood of $r \in R$, there exists a point $r' \in R$ such that, under iteration of $\widetilde{\phi}$, those two points do not lie in the same neighborhood any more*

Corollary 5. *We have* ***transitivity*** *: Given two neighborhoods U_1, U_2 in R, there is a point $r \in U_1$ such that r eventually lies in U_2 under iteration of $\widetilde{\phi}$*

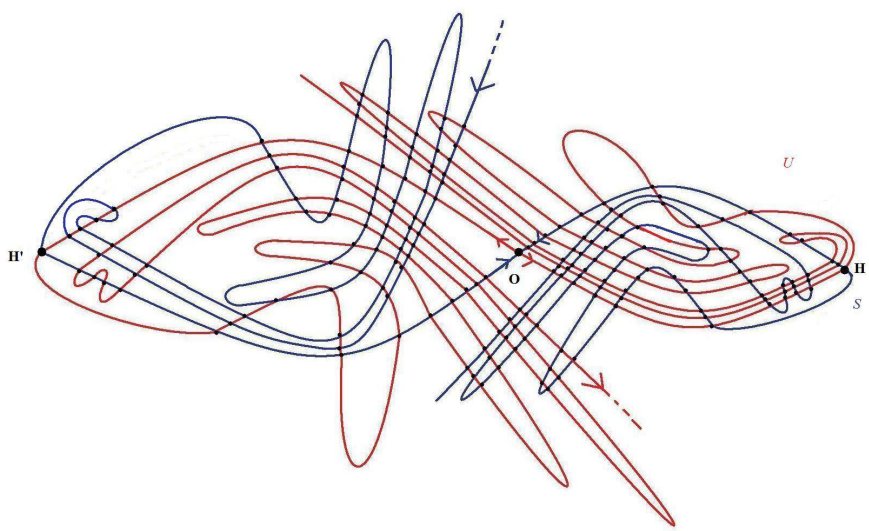
Corollary 6. *$\widetilde{\phi}$ is chaotic on R.*

Definition 24. *Given a diffeomorphism ϕ in a domain $D \subset \mathbb{R}^2$, a real-valued function f in D is an* ***integral*** *of ϕ if f is not constant, and $f(\phi(p)) = f(p)$*

Corollary 7. *ϕ does not possess a real analytic integral in R*

This completes the picture given at the beginning of the section:

as Dynamical Systems/Chaotic Behavior of Homoclinic Orbits.png



References

[1] R. Adler and B. Weiss. Similarity of automorphisms on the torus. *Mem. Amer. Math. Soc.*, 98, 1970.

[2] Douglas A. Lind and Brian Marcus. *An Introduction to Symbolic Dynamics and Coding*. Cambridge: Cambridge University Press, 1995.

[3] Peyam Ryan Tabrizian. Homoclinic orbits and chaotic behavior in classical mechanics. 2008.