An Introduction to Symbolic Dynamics

Robert Banks

January 9, 2018

Abstract

To find qualitative information about continuous objects, it is sometimes useful to consider their discrete analoges. For continuous dynamical systems, symbolic dynamics provide one such analog. The intent of this note is to provide some of the basic ideas and definitions behind symbolic dynamics as well as to give a few examples of how a symbolic representation of a system can be useful.

1 Introduction

Suppose we have a deterministic dynamical system defined for all time: a set X along with a mapping $\phi : X \times \mathbf{R} \to X$ which satisfies the condition that $\forall x \in X$, $\phi(x, 0) = x$. In this notation, we think of the real argument of ϕ as time, making $\phi(x, t)$ the location of an object starting at x after a time t. We can transform this system into a discrete sytem if we discretize both time and space.

The usual way to discretize time is to only look at the state of the system at some evenly spaced time intervals. For our purposes, we will do this by defining $T(x) = \phi(x, 1)$ and looking only at the values of $T(x)$. Note that the definition of $T(x)$ means that after n time steps, a point starting at x will have reached the point $T^n(x)$. Next, to discretize space, we partition X into a set of sets, $\{P_i\}$ so that $\cup_i P_i = X$ and, for $i \neq j, P_i \cap P_j = \emptyset$. While in theory this can be done arbitrarily, we usually restrict ourselves to using finitely many partition sets P_i. Our exact choice of partition will depend on the particular problems we wish to address.

Once we have discretized both space and time, we can begin describing our system symbolically. Let $x \in X$. This point, along with the map T, define a trajectory $\tau = (x, T(x), T^2(x), T^3(x), \cdots)$ through X which we can map to a sequence s by saying that if $\tau_i \in P_j$, then $s_i = j$. We think of this as a symbolic sequence, thinking of the indicies of the partitions as symbols. We call the set of symbols our *alphabet*, and any finite string of them form a *word*.

Example 1. *Let X be the unit interval mod 1, and define a map $\phi : X \times \mathbf{R} \to X$ by $\phi(x, t) = 2^t x \bmod 1$. Then $T(x) = \phi(x, 1) = 2x \bmod 1$, and the trajectory of x through X will be $\tau = (x, 2x \bmod 1, 4x \bmod 1, 8x \bmod 1, \cdots)$. Suppose we take* *our partition of X to be $P_0 = [0, \frac{1}{2})$ and $P_1 = [\frac{1}{2}, 1)$. In this case, if we take $x = \frac{3}{4}$, then $\tau = (\frac{3}{4}, \frac{1}{2}, 0, 0, \cdots)$. These are points in $(P_1, P_1, P_0, P_0, \cdots)$ respectively, so our symbolic sequence becomes $s = (1, 1, 0, 0, \cdots)$. A little examination shows that this map provides a base 2 expansion of x.*

Now so far, we've considered only the forward time evolution of our system, yet if our initial map ϕ is invertible, there is no reason to ignore the values of T^n when $n < 0$. To allow this (and to make our notation less cumbersome in general), we will henceforth drop the commas and parentheses and insert a period before our initial state. Thus, the symbolic sequence corresponding to $\tau = (\cdots, T^{-2}(x), T^{-1}(x), x, T(x), T^2(x), \cdots)$ will now be $s = \cdots s_{-2}s_{-1}.s_0 s_1 s_2 \cdots$, where s_i is the index of the partition containing $T^i(x)$. For the remainder of this note, we will state results using either one or two sided sequences. Generally speaking, the same results hold for the other sequence type as well. Also, we will now let our original system fade into the background and use the letter X to denote our symbolic systems, or *shift spaces*.

2 Shift Spaces

One benefit of translating our dynamical system into a symbolic setting is that advancing the system one step is particularly simple representation: moving our decimal point. This map, which we denote by σ, acts on a symbolic sequence $s = \cdots s_{-2}s_{-1}.s_0 s_1 s_2 \cdots$ by $\sigma s = \cdots s_{-1}s_0.s_1 s_2 s_3 \cdots$. The domain of this map is called a *shift space*. Now, from the above examples, we can distinguish two qualitatively different types of shift spaces. First, when we have an invertible map (and hence we have an infinite sequence of points before our decimal point), no information is lost by shifting, and we call σ a *full two-sided shift*, or *full A-shift* where A is our alphabet, or *full n-shift* if the cardinality of A is n. In the second case, when our map is not invertible, we have a *full one-sided shift* for which the domain of the function is the set of sequences with no numbers before the decimal point. Thus, for iterations of this map to be defined, we successively "forget" the first term in the sequence, saying that if $s = .s_0 s_1 s_2 s_3 \cdots$ then $\sigma s = .s_1 s_2 s_3 s_4 \cdots$.

Now, we should not generally expect that our systems will realize every possible symbolic trajectory. For instance, if symbol i corresponds to a trapping region in the original system, any symbolic sequence containing i will simply be i's forever after. Thus, we will generally be interested in some subspace of a full symbolic space. We call such subspaces *subshifts*, and require that they be closed under the shift map and with respect to a metric, which we now define.

Let u and v be two trajectories in a shift space. We then say that the distance between u and v is $\delta(u, v) = \frac{1}{2^m}$ where m is absolute value of the integer closest to zero for which $u_m \neq v_m$. Thus, for instance, if $u = \cdots 132.4513 \cdots$ and $v = \cdots 232.4513 \cdots$, then $\delta(u, v) = \frac{1}{2^2} = \frac{1}{4}$. This metric makes sense in that two trajectories are considered close if they correspond to paths starting in the same partition, coming from the same partitions, and going to the same partitions for some block of times near 0. Note, however, that does not always give the most natural definition of closeness in terms of the original system. In Example 1, for instance we could even have two points that are equivelent (say $0.49999\cdots$ and $0.50000\cdots$) that are nonetheless separated in this metric ($\delta(0.49999\cdots, 0.50000\cdots) = \frac{1}{2}$). This last example highlights the importance choosing an appropriate partition, and also suggests that the simple partitioning we are using may be inadequate for many real problems. The references below (such as [Lind and Marcus]) include more details on creating general partitions, but for our purposes, we will restrict ourselves to finite partitions of disjoint sets.

Now it turns out that, with these definitions, the structure of a subshift is simple to describe: a subshift is a set of all sequences not containing certain words. More formally, let F be any collection of finite strings of symbols in our alphabet. Then $X_F = \{\textit{all sequences not cantaining any word in } F\}$ is a subshift, and any subshift can be written in this form for a suitably chosen F. The susbshifts most commonly considered are those for which the set F is finite. These are called *subshifts of finite type* which are usually refered to as SFT's.

One benefit of SFTs is that, in SFTs, if one knows a large enough part of a trajectory, one can figure out all the ways the trajectory can continue. In particular suppose $s = s_0 s_1 \cdots s_n$ and F is our set of forbidden words. We then know that we can choose s_{n+1} to be anything that doesn't complete a forbidden word. Now, since F is finite, it has a longest word, of length, say, r. Thus, we can decide which values s_{n+1} can take by looking only at the last $r-1$ known entries of s.

Example 2. *Suppose that $A = \{0, 1\}$ and that $F = \{111\}$ ($r = 3$). If we have a partial sequence $s = .0001$, we only need to look at the last $r - 1 = 2$ entries to see what symbols can come next. Since these last symbls are 01 and no word of the form $1x$ or $01y$ for any $x, y \in A$ are forbidden, either 0 or 1 can come next. If, however, we add the one to get $.00011$,then our only next choice for a letter is 0.*

3 Higher Block Presentations

Now, while we can find our allowed next letters (or, in dynamic terms, allowed next states of our system) by looking at the previous $r-1$ letters (states), this can be cumbersome. It would be easier if we only needed to look back one symbol (state). While at first this seems overly optimistic, it is in fact possible. The idea that we translate our system into yet another symbolic system in which a block of the original sequence becomes a symbol in the new one, and the lengths of blocks correspond to the length of the longest forbidden word. The result of this procedure is known as a *higher block code.* Every SFT has such a *higher block presentation* in which the possibilities for the $(n+1)^{st}$ letter depends only on the n^{th}. Furthermore, these presentations are isomorphic or *conjugate* to the original system.

Example 3. *Let X_F be the symbolic system from Example 2 ($A = \{0, 1\}$,*

$F = \{111\}$, and $r = 3$). Since $r = 3$, we could map our system to another, Y_G in which each letter correspond to a three letter word in X_F. A natural way to do this would be to think of each three letter word in X_F as a base 2 number, convert it to a base 8 number, and consider this new number to be a symbol in Y_G. Then, for example, $110100001 \mapsto \overline{641}$. This done, all the forbidden words in Y_G are one or two letters. In particular, the forbidden words are $G = \{\overline{7}, \overline{16}, \overline{34}, \overline{36}, \overline{56}\}$. So, for example, if we see a $\overline{5}$ in our sequence, we know that the next character can be anything except $\overline{6}$.

One advantage of higher block presentations is that they allow us to use linear algebra to look at paths. To do this, we first think of our space as a graph with each of our partitions coresponding to a vertex and with a directed line segment connecting vertices i and j unless ij is a forbidden word. We can then construct an *adjacency matric* (or *transition matrix*), A, saying that $a_{ij} = 1$ if there is a directed edge from vertex i to vertex j and 0 otherwise.

Once we have constructed such an adjacency matrix, A, for our system, X, linear algebra provides us numerous results such as [Brin & Stuck]: the number of fixed points is the trace of A; the number of possible words of length $n+1$ starting at i and ending at j is the ij^{th} entry of A^n; and the number of periodic sequences in X with periods dividing n is the trace of A^n. We will mention some other uses for this graph representation below.

4 Topological Invariants

Given that one symbolic dynamical system allows for multiple representations, two natural question arise: First, how can we tell if two given different systems are, in fact, equivelent?; And second, if we translate a system from one form to another, what properties can we expect to be preserved? To begin answering these, we need a definition of what we mean for two systems to be equvelent. We start by calling a continuous map that commutes with the shift (a map ϕ such that $\phi \circ \sigma = \sigma \circ \phi$) a *homomorphism.* An invertible homomorphism is then known as an isomorphism or *conjugacy.* We consider systems equivelent if they are conjugate. Now unfortunately, our first question is difficult, in general, to answer directly. Thus, we will concentrate on the second here, and simply note that, if we have a property that is invariant under conjugation and two systems have different values of that property, then the systems cannot be conjugate.

One invariant is the number of n-periodic points in a system, where we define the set of period n points to be $Fix_n(X) = \{x \in X | f^n(x) = x\}$. Note that according to this definition, x is a period n point if the minimal period of x divides n. This information is sometimes encoded in the *Artin-Mazur zeta function*, which is defined by $\zeta_X(t) = \exp(\sum_{n=1}^{\infty} \frac{p_n(X)}{n} t^n)$, where $p_n(X)$ is the number of period n points in X. This function can also be expressed as the product $\zeta_X(t) = \prod_\gamma (1 - t^{|\gamma|})^{-1}$, where the product is taken over the periodic orbits, γ of X and $|\gamma|$ is the minimal period of γ. Looking at higher block representations above, we mentioned that if we can represent our system as a graph with adjacency matrix, A, then the number of paths of length n from the vertex i to itself is just the ii^{th} entry in A^n. Thus, the total number of period n fixed points is just the trace of A^n.

Another invariant is called *topological entropy.* This is essentially a measure of how fast the number of distinct trajectories in our system increases as the trajectory length increases. It is defined as $h(X) = lim_{n\to\infty} \frac{1}{n} \log N_n$ where N_n is the number of allowed words of length n in X. Since this measure is usually used only to compare systems, the base of the logarithm can be chosen for any given comparison. Choosing base 2 corresponds to measuring entropy in bits, while using base e corresponds to measuring in nats.

The computation of this invariant is yet one more reason to translate our system into a graph with adjacency matrix, A. An adjacency matrix has all of its entries greater than or equal to 0, so the Perron-Frobenius theorem tells that A has a positive eigenvalue, λ_A whose modulus is greater than or equal to the modulus of any other eigenvalue of A. This is a useful fact in entropy calculations because, with our system represented as a graph with adjacency matrix, A, $h(X) = log\lambda_A$.

Two last invariants that we will mention are the qualities of being *topologically transitive* or *topologically mixing.* A system is topologically transitive if any nonempty open set can be shifted to intersect any other nonempty open set, and topologically mixing if any sufficiently large shift of a nonempty open set intersects with another given open set. Formally, let U, and V be open sets in a shift space, X. The system is then topologically transitive if there is an $n > 0$ for which $\sigma^n(U) \cap V \neq \emptyset$ and is topologically mixing if there is an n_0 such that, if $n > n_0$, then $\sigma^n(U) \cap V \neq \emptyset$.

5 Ramsey Theory

In this section, we conclude with one final example of how translating a dynamical system into symbolic form can yield new information: that of Ramsey theory. Ramsey theory encompasses a range of results about what structures of a system cannot be destroyed by partitioning it. To be concrete, we consider one particular theorem by van der Waerden.

Van der Waerden's theorem states that if the integers are partitioned into a finite number of sets—$\mathbf{Z} = \cup_{i=1}^{n} P_i$—then for any k, there is an i such that P_i contains an arithmatic progression of length k.[1] To translate this into dynamical terms, we simply assign each interger m the partition corresponding to the symbol s_m of $s = \cdots s_{-2}s_{-1}.s_0s_1s_2\cdots$. We then interpret the result as follows: Let (X,T) be a dynamical system with $X = \cup_{i=1}^{n} P_i$, T the time 1 map, and take $x \in X$. Then, for any k, there is a single partition P_i and integers t_0 and m such that $T^t(x)$ is in P_i for $t = m, m+t_0, m+2t_0, \cdots, m+kt_0$. Thus van der Waerden's theorem guaruntees a certain type of regular recurance in a system.

[1] This formulation is equivelent to the original: Suppose $\mathbf{Z} = \cup_{i=1}^{n} P_i$ and choose $k \in \mathbf{N}$ Then there is an $M \in \mathbf{N}$ such that, for some i, the set $\{1, \cdots, M\} \cap P_i$ contains an arithmatic progression of length k.

Furthermore, these kinds of results can be used to show more general types of recurance. For instance, if a higher block code is being used, then the theorem shows that a sequence of states is garunteed to recur at regular intervals. If, on the other hand, a nonequal partition of time is used, then we will lose many of the features of symbolic dynamics, but will be able to find other kinds of recurant behavior.

Example 4. *Suppose* (X, T) *is a dynamical system,* $X = \cup_{i=1}^{n} P_i$*,* T *the time* 1 *map, and take* $x \in X$*. Define a sequence* $s = \cdots s_{-2}s_{-1}.s_0 s_1 s_2 \cdots$ *by* $s_j = i$ *if* $T^{j^3}(x) \in P_i$ *and choose an integer* k*. Then there is a partition* P_i *and are integers* j *and* t_0 *such that* $T^t(x)$ *is in* P_i *for the times* $t = j^3, (j+t_0)^3, (j+2t_0)^3, \cdots, (j+kt_0)^3$*.*

References

[1] Michael Brin and Garrett Stuck, *Introduction to Dynamical Systems.* Cambridge University Press, United Kingdom, 2002.

[2] Bruce Kitchens, *Symbolic Dynamics*, Springer, New York, 1991.

[3] Petr Kurka, *Topological and Symbolic Dynamics*, Société Mathématique de France, France, 2003.

[4] Douglas Lind and Brian Marcus, *An Introduction to Symbolic Dynamics and Coding*, Cambridge University Press, New York, 1995.

[5] Susan Williams, Editor, *Symbolic Dynamics and its Applications* American Mathematical Society, Rhode Island, 2002.