An Introduction to Geometric Topology and Dynamical Systems

Contents

CHAPTER 1

Introduction

This course includes topics from both dynamical systems and geometric and algebraic topology. Other references might include An Introduction of the Modern Theory of Dynamical Systems by Katok and Hasselblatt, Algebraic Topology by Allen Hatcher, and Introduction to Topology by Ganelin and Greene.

The aim of the course is not to develop the general theory of metric spaces or abstract topology in the most general setting. Most of the spaces which we deal with will be the nicest possible: Compact or closed subsets of $\mathbb{R}^n$. We will introduce dynamical and topological concepts concurrently, seeing that many of the ideas of topology and dynamics are inherently intertwined. The only prerequesite for a reading of these notes is a mathematical maturity obtained from, for instance, an abstract algebra or elementary real analysis course.

CHAPTER 2

Introductory Metric Space Topology

Dynamical Systems can take many contexts. In some cases, one wishes to consider the statistical properties of dynamical systems: the "average" value a system can take with respect to various quantities, such as temperature, velocity or height in a physical system. In some cases, we wish to restrict our attention to dynamical systems which are more accessible to modern methods, such as dynamical systems which are differentiable. In each case, different notions of equivalence are used. This chapter introduces the context on which our analysis of dynamical systems will take place: metric spaces.

2.1. Elementary Properties of Metric Spaces

The definitions and simple results of this section (or ones in a more narow or more general context) may be well-known to readers who have taken a course in topology or real analysis. Readers who have not encountered the topics in this section may wish to proceed at a slow pace. The following definition will be the one we use for the systems we consider.

Definition 2.1.1. A *metric space* is a pair (X, d), where X is a set and $d : X \times X \to [0, \infty)$ is a function satisfying for every $x, y, z \in X$:

(a) $d(x, y) = 0$ if and only if $x = y$
(b) $d(x, y) = d(y, x)$
(c) $d(x, y) + d(y, z) \geq d(x, z)$

Note that if (X, d) is a metric space and $Y \subset X$ is a subset of X, then $(Y, d|_{Y\times Y})$ is a metric space. We often abuse notation and simply write d instead of $d|_{Y\times Y}$. With this choice of d, Y is said to *inherit the ambient metric*, and this will often be assumed.

We give a few examples of metric spaces which will be central to our study. We leave the fact the metrics satisfy properties (a)-(c) as an exercise:

Example 2.1.1.

1. Take $X = \mathbb{R}$ or $\mathbb{C}$ and set $d(x, y) = |x - y|$.
2. Take $X = \mathbb{R}^n = \mathbb{R} \times \mathbb{R} \times \dots \mathbb{R}$, and set $d(x, y) = ||x - y||_2 = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$.
3. Take $X = \mathbb{R}^n$ and set for $p \in [1, \infty)$: $d_p(x, y) = ||x - y||_p = \left(\sum_{i=1}^{n} |x_i - y_i|^p\right)^{1/p}$
4. Take $X = \mathbb{R}^n$ and set $d_\infty(x, y) = \max\{|x_i - y_i| : i = 1, \dots, n\}$.
5. Take $X = \{z \in \mathbb{C} : |z| = 1\} \subset \mathbb{C}$ and use the ambient metric (equivalently, we divide $\mathbb{C}$ into real and imaginary axes, and say that $X = \left\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\right\}$). This space is denoted S^1 and is the round circle in $\mathbb{C}$.

We now present many definitions which lay the foundations of topology:

Definition 2.1.2. Let (X, d) and (Y, d') be two metric spaces:

(a) Given $\varepsilon > 0$ and $x \in X$, let $B_\varepsilon(x) = \{y \in X : d(x, y) < \varepsilon\}$. It is called the ε-ball around x.
(b) If $A \subset X$, let $B_\varepsilon(A) = \cup_{x\in A} B_\varepsilon(x)$. $B_\varepsilon(A)$ is called the *ball of radius ε around A*.
(c) A subset $U \subset X$ is said to be *open* if for every $x \in U$, there is some $\varepsilon > 0$ such that $B_\varepsilon(x) \subset U$. A subset $C \subset X$ is *closed* if $X \setminus C$ is an open set.
(d) A function $f : X \to Y$ is said to be *continuous* if $f^{-1}(V)$ is an open subset of X whenever V is an open subset of Y

(e) A sequence (x_n) of points in X is said to converge to x if for each $\varepsilon > 0$, there is an $N \in \mathbb{N}$ such that if $n \geq N$, $d(x_n, x) < \varepsilon$. We will often use the shorthand $x_n \to x$ to indicate that the sequence (x_n) converges to x.

EXAMPLE 2.1.2. The image below shows the balls $B_1(0) \subset \mathbb{R}^2$ when equipped with the norms d_p. When $p = 1$, the norm is a square whose sides intersect the axes at 45° angles. As p increases to 2, the ball expands to the round ball (it will have boundary S^1). As p increases to ∞, the unit balls actually "converge" to the unit ball of d_∞ (hence the notation d_∞), which is another square whose sides have length 2 and are parallel to the axes.

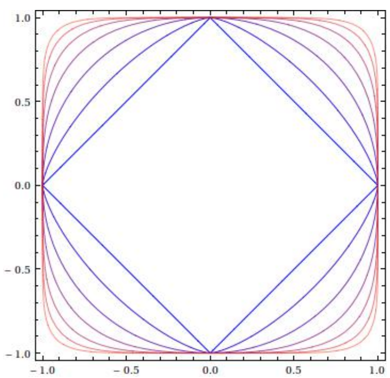

FIGURE 1. Unit balls of d_p for select values of p

DEFINITION 2.1.3. Let (X, d) be a metric space.

(a) Given a sequence (x_n) of points in X, a *subsequence* is a subset (x_{n_k}), with $n_1 < n_2 < \ldots$. It is considered a sequence in its own right by setting the k^{th} element of the sequence as x_{n_k}.
(b) Given $x \in X$ and $A \subset X$, x is a *limit point* or *accumulation point* of A if there is a sequence of distinct points (x_n) contained in A which converges to x
(c) A point x is said to be *isolated* if there is no sequence of distinct points converging to x. A space is *discrete* if every point is isolated.
(d) If $A \subset X$, define $\overline{A}$, called the closure of A, to be the union of A will its accumulation points
(e) A subset $A \subset X$ is dense if $\overline{A} = X$
(f) A function $h : X \to Y$ is said to be a *homeomorphism* if h is bijective, and both h and h^{-1} are continuous.
(g) A function $h : X \to Y$ is said to be an *isometry* if $d_Y(h(x_1), h(x_2)) = d_X(x_1, x_2)$ for all x_1 and x_2.

EXAMPLE 2.1.3. (a) Let $X = \mathbb{R}$, and $A = \mathbb{Z}$. Then the accumulation points of A is the empty set, since for any $n \in \mathbb{Z}$, $B_{1/2}(n) = \{n\}$. If we endow A with the subspace topology, A is discrete.
(b) Let $X = \mathbb{R}^2$, and $A = \mathbb{Z}^2$. Again, the accumulation points of A is the empty set. A is also discrete.
(c) Let $X = \mathbb{R}$ and $A = \{1/n : n \in \mathbb{Z}\}$. Then the set of accumulation points is $\{0\}$.
(d) Let $X = \mathbb{R}$ and $A = \mathbb{Q}$. Then the set of accumulation points is $\mathbb{R}$, since each point of $\mathbb{R}$ can be approached by distinct rationals.

Exercises.

1. Show that $B_\varepsilon(A)$ is an open set for every set $A \subset X$ and $\varepsilon > 0$
2. Show that for every $x \in X$, $\{x\}$ is a closed set
3. Show that x is isolated if and only if $\{x\}$ is an open set.

4. Show that $x_n \to x$ if and only if $d(x_n, x) \to 0$, where $d(x_n, x)$ is considered as a sequence in $(\mathbb{R}, d_1)$
5. Show that if (x_n) converges to both x and x', then $x = x'$
6. Show that the following are equivalent:
   (i) $f : X \to Y$ is continuous
   (ii) For each $x \in X$ and $\varepsilon > 0$, there exists a $\delta > 0$ such that if $d_X(x, x') < \delta$, then $d_Y(f(x), f(x')) < \varepsilon$
   (iii) If $x_n \to x$ in X, then $(f(x_n)) \to f(x)$ in Y
7. Prove or find a counterexample:
   (i) If f is an isometry, f is continuous.
   (ii) If f is an isometry, f is injective.
   (iii) If f is an isometry, f is a homeomorphism.
8. Show that if (x_n) is a sequence, the set of all limits of convergent subsequences of (x_n) can be expressed by:

$$\bigcap_{N \geq 0} \overline{\{x_n : n \geq N\}}$$

9. Show that $C \subset X$ is closed if and only if it contains its accumulation points. That is $C = \overline{C}$.
10. Show that for any $A \subset X$:

$$\overline{A} = \bigcap_{\substack{C \supset A \\ C \text{ closed}}} C = \bigcap_{n \in \mathbb{N}} B_{1/n}(A)$$

11. Let $X = [0, 1) \subset \mathbb{R}$ inherit the ambient metric and define $\theta : X \to S^1$ by $\theta(x) = \cos(2\pi x) + i \sin(2\pi x)$. Show that θ is continuous and bijective, but is not a homeomorphism.
12. Let $X = [0, 1)$ have the following metric:

$$d(x, y) = \min \{|x - y|, 1 - |x - y|\}$$

   Show that d is a metric and that with this metric, θ is a homeomorphism
13. Define a relation $(X, d_X) \sim (Y, d_Y)$ if and only if there exists a homeomorphism $h : X \to Y$. Show that $\sim$ is an equivalence relation. Homeomorphisms provide a natural way of classifying metric spaces, and this notion of equivalence will be used throughout the course.[1]
14. Show that $\mathrm{id} : (\mathbb{R}^n, d_p) \to (\mathbb{R}^n, d_q)$ is a homeomorphism for $p, q \in [1, \infty)$. In this way, we consider $\mathbb{R}^n$ with the distances d_p and d_q to be the same *topological* space, even though they differ as metric spaces. In general if d and d' are two metrics on X and $\mathrm{id} : (X, d) \to (X, d')$ is a homeomorphism, we say d and d' *induce the same topology.*
15. Define the distance $d_\angle : S^1 \times S^1 \to \mathbb{R}$ by $d_\angle(x, y) =$ the smaller of the two angles between the lines $\overline{0x}$ and $\overline{0y}$ (in the figure, $d_\angle(x, y) = \theta$). Show that $d_\angle$ is a metric and that d_2 and $d_\angle$ induce the same topology. [Hint: Use Exercise 12]
16. Let $Z = \{1/n : n \in \mathbb{N}\} \subset \mathbb{R}$, and consider a sequence (x_n) in X to be a map $S_x : Z \to X$ by $S_x(1/n) = x_n$. Show that x_n converges if and only if there exists a continuous map $\widetilde{S}_x : \overline{Z} \to X$ such that $\widetilde{S}_x|_Z = S_x$.

2.2. Compactness

We now introduce a fundamental notion in topology which can be thought of as a generalization of "finite." We will see that in most of the situations we will encounter, compactness means closed and bounded. It is a topological analog of finite volume.

[1] There is a logical problem associated with this exercise. While morally correct, it is rigorously taboo to consider the set of all metric spaces (for it contains the set of all discrete spaces, which contains the set of all sets). The authors will ignore this problem in the future as well, believing the intuition obtained from such exercises outweighs any objections from logicians.

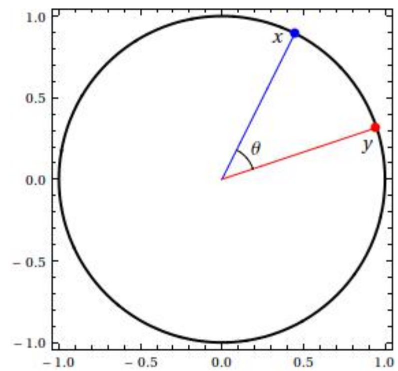

FIGURE 2. Calculating the $d_\angle$ distance between $x, y \in S^1$

DEFINITION 2.2.1. A metric space (X, d) is *compact* if for every sequence (x_n), there exists a subsequence (x_{n_k}) which converges. $K \subset X$ is compact if it is compact when given the ambient metric.

There are other notions of compactness, all of which are equivalent in metric spaces (our definition is sometimes called *sequential compactness*). Before we can prove the equivalence, we need to introduce a notion known as a Lebesgue number for a covering:

DEFINITION 2.2.2. Let X be a set in a topological space (it can be the space itself). An *open cover* of X is a collection of open sets $\mathcal{U} = \{U_\alpha\}$, where α runs over some indexing set (usually $\mathbb{N}$). If $\mathcal{U}$ is an open cover of a space X, δ is a *Lebesgue number* for $\mathcal{U}$ if for every $x \in X$, there is some $U \in \mathcal{U}$ such that $B_\delta(x) \subset U$.

We will exhibit an example of a cover with a Lebesgue number, and one without. Let $X = [0, 1]$ with the standard metric, and $\mathcal{U}_n = \left\{B_{(1+\varepsilon)/n}(k/n) : k = 0, 1, \ldots, n\right\}$. Then we claim that ε/n is a Lebesgue number for $\mathcal{U}_n$. To check this, suppose that $x \in [0, 1]$. Then there exists some k/n such that $|x - k/n| < 1/n$. Then if $d(x, y) < \varepsilon/n$, we know that $d(y, k/n) \leq d(y, x) + d(x, k/n) < \varepsilon/n + 1/n = (1 + \varepsilon)/n$. So $B_{\varepsilon/n}(x) \subset B_{(1+\varepsilon)/n}(k/n) \in \mathcal{U}_n$.

We will soon see that such a number exists for any open cover of a compact metric space. Let's first check that compactness is a requirement. We will construct an open cover of $X = (0, 3/4)$. Let $\mathcal{U} = \{B_{2^{-(n+1)}}(2^{-n}) : n \in \mathbb{N}\} = \left\{B_{1/4}(1/2), B_{1/8}(1/4), B_{1/16}(1/8), \ldots\right\}$. The union of the sets in $\mathcal{U}$ is clearly open in $\mathbb{R}$, contains arbitrarily small numbers, and since the left endpoint of one of the intervals is the center of the next, it is an interval. The rightmost endpoint is $3/4$, so it covers $(0, 3/4)$.

Claim. $\mathcal{U}$ has no Lebesgue number.

PROOF. Assume, to reach a contradiction, that δ is a Lebesgue number for $\mathcal{U}$. Choose n such that $\delta > 2^{-n}$. Then $B_\delta(2^{-(n-1)}) \supset B_{2^{-n}}(2^{-(n-1)})$, and the containment is strict. The only set in $\mathcal{U}$ that contains $2^{-(n-1)}$ is exactly the right hand side of the strict containment. Thus, we have found a point $x = 2^{-(n-1)}$ such that $B_\delta(x)$ is not contained in any set in $\mathcal{U}$, and $\mathcal{U}$ has no Lebesgue number.

□

LEMMA 2.2.3 (Lebesgue number lemma). *If X is compact, every open cover of X has a Lebesgue number.*

PROOF. We will prove by contradiction. If $\mathcal{U}$ was an open cover without a Lebesgue number, then there would exist points x_n such that $B_{1/n}(x_n)$ was not contained in any $U \in \mathcal{U}$. By assumption, there is a subsequence x_{n_k} which converges to some point x. Since $\mathcal{U}$ covers X, there exists some ε such that $B_\varepsilon(x) \subset U$ for some $U \in \mathcal{U}$. Since $x_{n_k} \to x$, there exists some N such that $x_{n_k} \in B_{\varepsilon/2}(x)$ for every $n_k \geq N$. But choosing $m = n_k > 2/\varepsilon$, we conclude that $x_m \in B_\varepsilon(x)$, which implies $B_{1/m}(x_m) \subset B_{\varepsilon/2}(x_m) \subset B_\varepsilon(x) \subset U$. This contradicts our choice of x_m. □

We leave it as an exercise to the reader to see where the proof breaks down for the cover $\mathcal{U}$ of $(0, 3/4)$

THEOREM 2.2.4. *The following are equivalent:*

(i) X is compact
(ii) If $A \subset X$ has infinitely many elements, A has an accumulation point
(iii) If U_i are a collection of open sets of X such that $X = \cup_{i=1}^\infty U_i$ (such a collection is called an open cover of X), then there are finitely many indices $\{i_1, \ldots, i_N\}$ such that $X = \cup_{i=1}^N U_{\alpha_i}$

PROOF.
(i) $\Leftrightarrow$ (ii)
Let A have infinitely many elements, so that there are distinct elements (x_n). Because X is compact, there is a convergent subsequence which lies in A, consisting of distinct elements. That is, A has a limit point.

Let x_n be a sequence in X. If $A = \{x_n\}$ is finite, then one of the points in A must occur infinitely often in the sequence x_n, and is hence the limit of a constant subsequence. If A is infinite, let x be an accumulation point. Choose a subsequence in the following way. Let $x_{n_1} = x_1$. Then use the inductive procedure by letting x_{n_k} be chosen so that $d(x_{n_k}, x) < d(x_{n_{k-1}}, x)/2$. One can inductively verify that $d(x_{n_k}, x) < d(x_1, x)/2^{k-1} \to 0$, and hence $x_{n_k} \to x$ as $k \to \infty$.

(i) $\implies$ (iii)
Let $\mathcal{U}$ be an open cover of X. By Lemma 2.2.3, X has a Lebesgue number δ. Assume, for the purpose of contradiction, that there was no finite subcover of $\mathcal{U}$. Then $\mathcal{U}$ must be infinite and has at least countable cardinality, so we may choose open subsets $\mathcal{U} \supset \{U_1, U_2, \ldots, U_n, \ldots\}$. We can insist that $U_n \not\subset \bigcup_{i=1}^{n-1} U_i$, since if this were the case for every open set U_n, the collection $\{U_1, \ldots, U_{n-1}\}$ would cover X, which we have assumed is not possible. So can insist that there exists some $x_n \in U_n$ such that $x_n \notin U_i$ for $i < n$, and $B_\delta(x_n) \subset U_n$ (the second condition can be guaranteed by picking x_n, and then possibly switching U_n to the appropriate open set guaranteed by the Lebesgue number lemma).

We assumed that X is compact, so we can conclude that x_n converges to some x. But then $d(x_n, x) \to 0$, so there exists N such that if $n \geq N$, $d(x_n, x) < \delta/2$. So if $m, n \geq N$:

$$d(x_m, x_n) \leq d(x_m, x) + d(x, x_n) < \delta$$

But this is a contradicts our construction, because then:

$$x_{N+1} \in B_\delta(x_N) \subset U_N$$

(iii) $\implies$ (i)
We prove the contrapositive. That is, we assume there exists a sequence (x_n) for which every subsequence does not converge, and attempt to show that we can find a collection of open sets U_n which cover X but has no finite subcover. We claim that for each x_i, there exists some $\varepsilon_i > 0$ such that $x_j \notin B_{\varepsilon_i}(x_i)$ for every $j \neq i$. If not, there would exist n_k such that $x_{n_k} \in B_{2^{-k}}(x_i)$, and hence x_{n_k} would be a convergent subsequence. Now let $U_i = B_{\varepsilon_i}(x_i)$ and $U_0 = X \setminus \{x_1, x_2, \ldots\}$. Then $X = \cup_{i=0}^\infty U_i$, but $x_i \in U_j$ if and only if $i = j$, so no finite subcollection covers X. □

PROPOSITION 2.2.5. *Any interval $[a, b]$ is compact*

Proof. We proceed by contradiction, using the definition of compactness by open covers. Assume that there is an open cover $\mathcal{U}$ of $[a, b]$ with no finite subcover. Set $I_0 = [a, b]$ and divide I_0 into subintervals $I_- = [a, (a+b)/2]$ and $I_+ = [(a+b)/2, b]$. Then $\mathcal{U}$ also covers I_- and I_+. If there are finite subcovers $\mathcal{U}_1$ of I_- and $\mathcal{U}_2$ of I_+, then $\mathcal{U}_1 \cup \mathcal{U}_2$ is a finite subcover of $[a, b] = I_- \cup I_+$, so we have that either I_- or I_+ has no finite subcover. Let I_1 denote one of the intervals which has no finite subcover. We can iterate this process, dividing the interval I_n into two equal subintervals. We get a sequence of intervals satisfying:

(1) $\text{length}(I_i) = 2^{-n}(b-a)$
(2) $\mathcal{U}$ covers I_i and has no finite subcover
(3) $I_0 \supset I_1 \supset \cdots \supset I_n \supset I_{n+1} \supset \ldots$

The nested interval property of $\mathbb{R}$ tells us that any sequence of intervals $I_n \supset I_{n+1}$, such that $\text{length}(I_n) \to 0$, $\bigcap_{n=1}^{\infty} I_n$ consists of exactly one point. The open cover U must cover this point, as it is in each of the I_n and $\mathcal{U}$ covers each I_n. Then $x \in U \in \mathcal{U}$ for some U, and since U is open, there exists some $\varepsilon > 0$ such that $B_\varepsilon(x) \subset U$. Choose n large enough so that $2^{-n} < \varepsilon/(b-a)$. Since $x \in I_n$ and $\text{length}(I_n) < (b-a)2^{-n}$, for every $y \in I_n$, $d(x, y) < 2^{-n} < \varepsilon$. Thus, $I_n \subset B_\varepsilon(x) \subset U$. Thus, $\{U\} \subset \mathcal{U}$ is a finite cover of I_n. This is a contradiction of our construction, as I_n cannot have a finite subcover.

□

Definition 2.2.6. Let X and Y be metric spaces, and $X \times Y = \{(x, y) : x \in X, y \in Y\}$ be the product of sets. Define the *product distance* by:

$$\hat{d}((x_1, y_1), (x_2, y_2)) = \max\{d_X(x_1, x_2), d_Y(y_1, y_2)\}$$

Proposition 2.2.7. *A sequence (x_n, y_n) in $X \times Y$ converges to (x, y) if and only if $x_n \to x$ and $y_n \to y$ in X and Y, respectively*

Proof. First, suppose that $(x_n, y_n) \to (x, y)$. Then $d(x_n, x) \leq \max\{d(x_n, x), d(y_n, y)\} = \hat{d}((x_n, y_n), (x, y)) \to 0$. Thus, $d(x_n, x) \to 0$, so $x_n \to x$. Similarly, $y_n \to y$.

Now, suppose that $x_n \to x$ and $y_n \to y$. Let $\varepsilon > 0$, and choose N_1 and N_2 such that if $n \geq N_1$, $d(x_n, x) < \varepsilon$ and if $n \geq N_2$, $d(y_n, y) < \varepsilon$. But then if $N = \max N_1, N_2$ and $n \geq N$, $\hat{d}((x_n, y_n), (x, y)) = \max\{d(x_n, x), d(y_n, y)\} < \varepsilon$. That is $(x_n, y_n) \to (x, y)$ □

Proposition 2.2.8. *Let X and Y be compact metric spaces.*

(1) If $Z \subset X$ is a closed subset, then Z is compact
(2) $X \times Y$ is compact

Proof.

(1) We use the definition of compactness by open covers. Suppose that $\mathcal{U}$ is an open cover of Z. Since Z is closed, $V = X \setminus Z$ is open. But then $\mathcal{U} \cup \{V\}$ is an open cover of X, which by assumption, must have a finite subcover $\mathcal{U}_0 \cup \{V\}$. But since $V \cap Z = \emptyset$, $\mathcal{U}_0$ covers Z. We have shown that every open cover of Z has a finite subcover, that is Z is compact.
(2) We use the definition of compactness by sequences. Let (x_n, y_n) be any sequence in $X \times Y$. Then since X is compact, there is a subsequence x_{n_k} converging to some $x \in X$. Similarly, since Y is compact, there exists a subsequence $y_{n_{k_l}}$ of y_{n_k} which converges to a point y. But then $(x_{n_{k_l}}, y_{n_{k_l}}) \to (x, y)$, so (x_n, y_n) contains a convergent subsequence, and $X \times Y$ is compact.

□

Theorem 2.2.9 (Heine-Borel). *If $X \subset \mathbb{R}^n$ is a metric space which inherits the ambient metric, X is compact if and only if X is closed and bounded.*

Proof. Since X is bounded, it must be contained in a set of the form $B = [-T, T]^n$. But by the above propositions this set is compact. Since X is a closed subset of the compact set B, X is compact. □

Exercises. Let (X, d_X) and (Y, d_Y) be metric spaces.

1. Show that compactness is a topological invariant. That is, show that if X is compact and $f : X \to Y$ is a homeomorphism, then Y is compact.
2. Let X be a compact metric space. Show that if $K \subset X$, K is compact if and only if K is closed
3. Show that if $f : X \to Y$ is continuous, and $K \subset X$ is compact, then $f(K)$ is compact
4. Show that if $X = [0, 1)$ is given the ambient metric (as in Exercise 11 of the previous section), X is not compact. Furthermore, show that if it is given the metric described in Exercise 12 of the previous section, X is compact.
5. Show that if X is a compact metric space with at most countably many points, then X has finitely many points.
6. A simple, closed curve in $\mathbb{R}^2$ is a continuous map $\gamma : [0, 1] \to \mathbb{R}^2$ which is injective on $[0, 1)$ and has $\gamma(0) = \gamma(1)$. The purpose of this exercise is to show that for any simple, closed curve $\gamma : [0, 1] \to \mathbb{R}^2$, $\gamma([0, 1])$ is homeomorphic to S^1.
   (i) Show that if X is compact, and $h : X \to Y$ is continuous and bijective, then h is a homeomorphism
   (ii) Let $\gamma_0 : [0, 1) \to \mathbb{R}^2$ be the restriction of γ to this set. Show that γ_0 is continuous with respect to the metric of Exercise 12.
   (iii) Conclude from part (i) and Exercise 4 that γ_0 is a homeomorphism onto its image.
   (iv) Conclude from part (iii), Exercises 12 and 13, that S^1 and $\gamma([0, 1])$ are homeomorphic.

2.3. Completeness

Compactness is related to another notion called completeness. Metric spaces which are complete have the convenient property that if a sequence *should* converge to something, it must. We make this precise using the following definition.

Definition 2.3.1. Let (X, d) be a metric space. A sequence (x_n) is *Cauchy* if for every $\varepsilon > 0$, there exists some $N \geq 0$ such that if $m, n \geq N$, $d(x_n, x_m) < \varepsilon$. X is said to be *complete* if every Cauchy sequence converges.

We list a few examples which are assumed to be known.

Example 2.3.1.

(1) $\mathbb{R}$ equipped with the standard metric is complete
(2) $(0, 1)$ is *not* complete. The sequence $\frac{1}{n}$ is Cauchy but does not converge in $(0, 1)$
(3) $\mathbb{Q} \subset \mathbb{R}$ inheriting the standard metric is *not complete*
(4) Equip $(0, \pi)$ with the metric $\hat{d}(x, y) = |\cot(x) - \cot(y)|$. We leave as an exercise to the reader that this space is complete [*Hint*: Show that the topology induced by $\hat{d}$ is the same as that of the standard metric, but that any sequence approaching 0 or π in the standard metric is not Cauchy with respect to $\hat{d}$]

As part 4 demonsrates, completeness is *not* a topological invariant, because $(0, 1)$ and $\mathbb{R}$ are homeomorphic. Rather, it is a *metric* invariant.

2.3.1. Baire Spaces.

Here is a natural question: If (X, d) is a metric space, does there exist a metric d' on X which induces the same topology and such that (X, d') is complete (as was the case in part 4)?

The general answer is no, but can we find a counterexample? It turns out that $\mathbb{Q}$ equipped with the standard metric works nicely. We need a powerful theorem from metric space topology to verify this.

Definition 2.3.2. A subset $A \subset X$ is a G_δ set if there exist open sets $U_i \subset X$ such that $A = \bigcap_{i=1}^\infty U_i$. $B \subset X$ is an F_σ set $B = X \setminus A$, where A is a G_δ set.

Definition 2.3.3. X is said to be a *Baire space* if whenever U_i are open and dense for each i, then $\bigcap_{i=1}^\infty U_i$ is a dense G_δ set.

Example 2.3.2. $\mathbb{R} \setminus \mathbb{Q}$ is a dense G_δ set, since $R \setminus \mathbb{Q} = \mathbb{R} \setminus \cup \{q_i\} = \cap(\mathbb{R} \setminus \{q_i\})$

Note that because we can formulate the property of being a Baire space using open sets and topological invariants, then this is a topological property (ie, it is preserved by homeomorphisms), using the following Lemma:

Lemma 2.3.4. *Let X be a metric space. $A \subset X$ is dense if and only if for every nonempty open set $U \subset X$, $A \cap U \neq \emptyset$*

Proof. First, suppose that A is dense. Then if U is a nonempty open set, there exists some $x \in U$. But then since U is open, there exists $\varepsilon > 0$ such that $B_\varepsilon(x) \subset U$. Since A is dense, we either have $x \in A$, or x is an accumulation point of A. If $x \in A$, then $x \in A \cap U$. If x is an accumulation point, then there exists a sequence (x_n) in A approaching x. In particular, there exists some $x_N \in B_\varepsilon(x) \subset U$. So $x_N \in U \cap A$.

Now suppose that we have the intersection property. Let $x \in X$. If $x \in A$, we're done. So we need to show that if $x \notin A$, x is an accumulation point of A. By the above property, since $B_{1/n}(x) \cap A \neq \emptyset$, there exists some $x_n \in B_{1/n}(x) \cap A$. We know that $x_n \neq x$ since $x \notin A$. Then $x_n \to x$. □

We can now state the theorem:

Theorem 2.3.5 (Baire Category). *If (X, d) is a complete metric space, then X is a Baire space.*

Let us see how this helps with the case of $\mathbb{Q}$, pushing the proof until later. We claim that $\mathbb{Q}$ is not a Baire space. Then the lack of a distance making $\mathbb{Q}$ complete will follow from:

Corollary 2.3.6. *If (X, d) is not a Baire space, then there does not exist a metric d' on X (inducing the same topology of X) for which (X, d) is complete*

Proof. Suppose there existed such a metric. Then $\mathrm{id} : (X, d') \to (X, d)$ is a homeomorphism. Since (X, d) is not a Baire space, there exist open dense sets $U_i \subset X$ such that $\bigcap_{i=1}^\infty U_i$ is not dense. But by Lemma 2.3.4, open dense sets in (X, d') are still open and dense in (X, d). Hence there cannot exist such a metric, since it would violate the Baire category theorem. □

Let us verify $\mathbb{Q}$ is not a Baire space. To do so, let's construct open dense sets U_i. We know that $\mathbb{Q}$ is countable. That is $\mathbb{Q} = \{q_1, q_2, \ldots\}$. Then let $U_i = \mathbb{Q} \setminus \{q_i\}$. Then U_i is open and dense, and by construction $\cap_{i=1}^\infty U_i = \emptyset$, which is obviously not dense in $\mathbb{Q}$.

Proof of Theorem 2.3.5. Let U_i be a collection of open, dense subsets of X. We want to show that $\cap_{i=1}^\infty U_i$ is dense. By Lemma 2.3.4, this holds if and only if for any open set $V \subset X$, $\cap_{i=1}^\infty U_i \cap V \neq \emptyset$. Now, because U_1 is open and dense, $U_1 \cap V$ is open and nonempty. Then there exists x_1 and $r_1 < 1$ such that $\overline{B_{r_1}(x_1)} \subset U_1 \cap W$. Now U_2 is open and dense, so $U_2 \cap B_{r_1}(x_1)$ is open and nonempty. Thus there exists a ball $\overline{B_{r_2}(x_2)} \subset U_2 \cap B_{r_1}(x_1)$ and $r_2 < 1/2$. Continuing in this way, we get a sequence of balls $B_{r_i}(x_i) \subset U_i \cap B_{r_{i-1}}(x_{i-1})$ and $r_i < 2^{-i}$. In particular:

$$\ldots B_{r_i}(x_i) \subset \overline{B_{r_i}(x_i)} \subset B_{r_{i-1}}(x_{i-1}) \subset \overline{B_{r_{i-1}}(x_{i-1})} \subset B_{r_{i-2}}(x_{i-2}) \cdots \subset V$$

We claim that (x_n) is a Cauchy sequence, which converges to a point in the intersection of the U_i. Take $\delta > 0$, and choose N such that $2^N < \delta$. Then if $m \geq n \geq N$, $x_m \in B_{r_n}(x_n)$, and $d(x_m, x_n) < r_n < 2^{-n} < 2^{-N} < \delta$. Thus (x_n) is Cauchy.

Thus $x_n \to y$ for some y. We claim that $y \in U_i$ for every i and $y \in V$. This follows because we chose the *closure* of $B_{r_i}(x_i)$ to be a subset of U_i, so any convergent sequence contained in it will converge to a point of $\overline{B_{r_i}(x_i)}$, and hence a point of U_i. □

The following theorem allows us to recognize that many of the spaces we work with in dynamical systems are in fact complete:

THEOREM 2.3.7. *If (X, d) is a compact metric space, it is complete*

PROOF. Let (x_n) be a Cauchy sequence in X. Since X is compact, it has a convergent subsequence $x_{n_k} \to x$. We claim that $x_n \to x$. To see this, let $\varepsilon > 0$. Choose N such that if $n_k \geq N$, then $d(x_{n_k}, x) < \varepsilon/2$, and that if $m, n \geq N$, that $d(x_m, x_n) < \varepsilon/2$. Without loss of generality, we can assume that $N = n_k$ for some k. Then if $n \geq N$:

$$d(x_n, x) \leq d(x_n, x_N) + d(x_N, x) < \varepsilon/2 + \varepsilon/2 = \varepsilon$$

□

2.3.2. Function Spaces. We give another example of a complete space, this time in an infinite-dimensional setting.

DEFINITION 2.3.8. Let X be a metric space, Y be a compact metric space, and $C(X, Y)$ be the set of continuous functions from X to Y. Define the following metric on $C(X, Y)$:

$$d(f, g) = \sup_{x \in X} d_Y(f(x), g(x))$$

Note that the metric is well-defined because Y is compact and hence the supremum is finite.

THEOREM 2.3.9. *If X is compact and Y is complete, then $C(X, Y)$ is complete.*

PROOF. We need to show that if we have a sequence of functions $f_n \in C(X, Y)$ which is Cauchy, then f_n must converge. We use the following function as a candidate for the limit:

$$f(x) = \lim_{n \to \infty} f_n(x)$$

First, we claim that f must be well-defined, since if $d(f_n, f_m) < \varepsilon$, then for every x, $d_Y(f_n(x), f_m(x)) < \varepsilon$, so each sequence $f_n(x)$ must be Cauchy. Since Y is complete the sequence must converge.

Note that while we don't know that f is continuous yet, we can still make sense of $d(f_n, f)$. We claim that $d(f_n, f) \to 0$ as $n \to \infty$. Indeed, since f_n is Cauchy, for each $\varepsilon > 0$ there exists N such that if $m, n \geq N$, $d(f_n, f_m) < \varepsilon$. Then if $n \geq N$:

$$d(f_n, f) = \sup_{x \in X} d(f_n(x), f(x)) = \sup_{x \in X} d\left(f_n(x), \lim_{m \to \infty} f_m(x)\right) = \sup_{x \in X} \lim_{m \to \infty} d(f_n(x), f_m(x)) \leq \varepsilon$$

Now we show (uniform) continuity of f. Let $\varepsilon > 0$. Then there exists $n \in \mathbb{N}$ such that $d(f_n, f) < \varepsilon/3$ and a $\delta > 0$ such that if $d_X(x_1, x_2) < \delta$, then $d_Y(f_n(x_1), f_n(x_2)) < \varepsilon/3$. So we compute:

$$d_Y(f(x_1), f(x_2)) \leq d_Y(f(x_1), f_n(x_1)) + d_Y(f_n(x_1), f_n(x_2)) + d_Y(f_n(x_2), f(x_2)) < 3 \cdot \frac{\varepsilon}{3} = \varepsilon$$

So f is uniformly continuous.

□

2.3.3. Completions. If X is a metric space, it always naturally sits inside another complete metric space $\widehat{X}$, called the completion of X.

DEFINITION 2.3.10. If X is a metric space, Y is a *completion* of X if Y is complete and there exists an isometry $f : X \to Y$ such that $\overline{f(X)} = Y$.

EXAMPLE 2.3.3. Consider the metric space $\mathbb{Q}$. We have seen that $\mathbb{Q}$ does not admit a metric equivalent to the standard one making it complete. But the inclusion map $i : \mathbb{Q} \to \mathbb{R}$ shows that $\mathbb{R}$ is a completion of $\mathbb{Q}$, which adds in all of the limits of Cauchy sequences in $\mathbb{Q}$.

Completions are *not* topological invariants (rather, they invariants of the metric on the topological space). That is, there are two homeomorphic topological spaces which have different completions. For example, $\mathbb{R}$ is complete, so $\widehat{\mathbb{R}} = \mathbb{R}$. But $\mathbb{R}$ is homeomorphic to $(0, 1)$ with the homeomorhpism $f : (0, 1) \to \mathbb{R}$ given by $f(x) = x/(1 - x^2)$. The completion of $(0, 1)$ is $[0, 1]$. But $[0, 1]$ and $\mathbb{R}$ are not homeomorphic, since $[0, 1]$ is compact and $\mathbb{R}$ is not.

The following Proposition shows that completions are unique up to isometry:

PROPOSITION 2.3.11. *If Y_1 and Y_2 are two completions of X, with associated maps $f_i : X \to Y_i$ then there exists a unique bijective isometry $F : Y_1 \to Y_2$ such that $f_2 = F \circ f_1$.*

PROOF. Note that since $f_i(X)$ is dense in Y_i for $i = 1, 2$, any point in $y \in Y_1$ can be written as the limit of a sequence $f_1(x_n) \to y$. Then we claim that we can define $F(y) = \lim_{n\to\infty} f_2(x_n)$.

We need to verify that F is well-defined. To see well-definedness, we need to verify that $f_2(x_n)$ converges and that the limit is independent of which sequence x_n we chose to converge to y. First, note that since $f_1(x_n)$ converges, it is Cauchy. Also,

$$d_{Y_2}(f_2(x_m), f_2(x_n)) = d_X(x_m, x_n) = d_{Y_1}(f_1(x_m), f_1(x_n))$$

since the f_i are isometries. Hence $f_2(x_n)$ is a Cauchy sequence and must converge since Y_2 is complete.

Now, suppose that x_n and z_n are two sequences such that $f_1(x_n)$ and $f_1(z_n)$ both converge to y. Then:

$$d_{Y_2}\left(\lim_{n\to\infty} f_2(x_n), \lim_{n\to\infty} f_1(z_n)\right) = \lim_{n\to\infty} d_{Y_2}(f_2(x_n), f_2(z_n)) = \lim_{n\to\infty} d_{Y_1}(f_1(x_n), f_2(z_n)) = 0$$

So the limits must coincide, and F is well-defined. We leave to the reader to verify that F is an isometry and satisfies $f_2 = F \circ f_1$.

To see that F is bijective, we construct an inverse. Note that we can do this procedure swapping the roles of Y_1 and Y_2 to get an isometry $G : Y_2 \to Y_1$. Then $G \circ F : Y_1 \to Y_1$ satisfies $G \circ F \circ f_1 = G \circ f_2 = f_1$. That is, $G \circ F|_{f_1(X)}$ is the identity on $f_1(X)$. Therefore, $G \circ F = \mathrm{id}$ by continuity and density of $f_1(X)$. Thus, F has an inverse on the left. A similar argument shows that G is a right inverse as well.

□

We give a construction for the completion of a metric space X using *Busemann functions.* In the exercises, we outline another construction using Cauchy sequences.

Let $B(X, \mathbb{R})$ be the set of all bounded (not necessarily continuous) functions from X to $\mathbb{R}$. We use the same distance on $B(X, \mathbb{R})$ that we do for $C(X, \mathbb{R})$:

$$d(f, g) = \sup_{x\in X} |f(x) - g(x)|$$

The supremum must exist since f and g are bounded.

LEMMA 2.3.12. *$B(X, \mathbb{R})$ is complete*

The proof that $B(X, \mathbb{R})$ is complete is identical to that of Theorem 2.3.9, but we no longer need to check continuity of the limit. We therefore refer the reader to this proof and omit the repetition.

Fix $x_0 \in X$. If $x \in X$, let $\delta_x \in B(X, \mathbb{R})$ be the function $\delta_x(y) = d(x, y) - d(y, x_0)$. We fist claim that δ_x is bounded. Indeed, $\delta_x(y) = d(x, y) - d(y, x_0) \le d(x, x_0)$ for every y.

Let us compute $d(\delta_x, \delta_z) = \sup_{y\in X} |d(x, y) - d(y, z)|$. Note that the supremum must be at least $d(x, z)$, since if we take $y = z$, this is exactly what we get. But we also know that $d(x, y) - d(y, z) \le$

$d(x,z)$ by the triangle inequality, so $d(\delta_x,\delta_y)=d(x,y)$, and the map $f(x)=\delta_x$ is an isometry. Then let $\widehat{X}=\overline{f(X)}$. We get that $\widehat{X}$ is a completion of X given Exercise 1.

Exercises.

1. Show that if X is a complete metric space and $A\subset X$ is a closed subset, then A is complete when given the ambient metric.
2. Show that if X is a complete metric space with countably many points, then X has an isolated point. [*Hint*: First show that $U_x=X\setminus\{x\}$ is dense for every x if X has no isolated points]
3. Prove or find a counterexample: If X is a complete metric space with countably many points, then X is discrete.
4. Prove or find a counterexample: If X is a discrete space, X is complete.
5. Prove or find a counterexample: If X is a complete, bounded metric space, then X is compact.
6. A set $A\subset X$ is said to be *nowhere dense* if for every ball $B_\varepsilon(x)\subset X$, $\overline{A\cap B_\varepsilon(x)}\neq\overline{B_\varepsilon(x)}$. Show that a set A is F_σ (see definition 2.3.2) if and only if A is a countable union of nowhere dense sets.
7. In this sequence of exercises, we give an alternate construction for the completion of a metric space X. Let Cauchy(X) denote the set of Cauchy sequences in the metric space X, and say that $(x_n)\sim(y_n)$ are equivalent if the interpolating sequence z_n defined by $z_{2n}=x_n$ and $z_{2n-1}=y_n$ is Cauchy. Then let $\widehat{X}=\text{Cauchy}(X)/\sim$ be the set of equivalence classes, and $[x_n]$ denote the equivalence class of the sequence (x_n).
   (a) Define a function d' on Cauchy$(X)\times$Cauchy(X) by $d'((x_n),(y_n))=\lim_{n\to\infty}d(x_n,y_n)$. Show that d' is well-defined, symmetric and satisfies the triangle inequality.
   (b) Show that $d_{\widehat{X}}([x_n],[y_n])=d'((x_n),(y_n))$ is a well-defined metric on $\widehat{X}$.
   (c) Show that if $x\in X$ and $\overline{x}_n=x$ is the sequence which is constantly x, then the map $f(x)=[\overline{x}]$ is an isometry and $f(X)$ is dense in $\widehat{X}$.
   (d) Show that if $A\subset Y$ is a dense subset of a metric space Y, then Y is complete if and only if every Cauchy sequence in A converges to a point in Y
   (e) Show that $\widehat{X}$ is a complete metric space
8. * In this sequence of exercises, we will show that the set of continuous functions $f:[0,1]\to\mathbb{R}$ which are not differentiable at any point are dense.The Anti-Hunger Gam
   (a) Show that if $f\in C([0,1],\mathbb{R})$, then the functions defined by:

$$f_n(x)=f(k/n)+(x-k/n)(f((k+1)/n)-f(k/n))\text{ when }x\in[k/n,(k+1)/n]$$

   satisfy $f_n\to f$.
   (b) Let $U_{m,n}=\left\{g\in C([0,1],\mathbb{R}):\forall x\in[0,1]\exists x'\in B_{1/m}(x)\text{ such that }d(g(x),g(x'))>nd(x,x')\right\}$. Show that $U_{m,n}$ is dense.
   [*Hint*: Given $f\in C([0,1],\mathbb{R})$, construct an element of $U_{m,n}$ arbitrarily close to f by first choosing some f_n from part (a) that approximates f well enough. Then add a function of the form:

$$g(x)=\begin{cases}\lambda(x-k/m) & x\in[k/m,(4k+1)/4m]\\ \lambda\left(1/(2m)-(x-k/m)\right) & x\in[(4k+1)/4m,(4k+3)/4m]\\ \lambda\left(-1/(2m)+(x-k/m)\right) & x\in[(4k+3)/4m,(k+1)/m]\end{cases}$$

   where m is a sufficiently large multiple of n allowing λ to dominate the slopes of f_n]
   (c) Show that $U_{m,n}$ is an open subset of $C([0,1],\mathbb{R})$.
   (d) Show that the set of functions which are not differentiable at any point are a dense G_δ subset of $C([0,1],\mathbb{R})$

2.4. Connectedness

In addition to completeness and compactness, a third property of metric spaces plays a crucial role in understanding metric space topology, as well as a tool in understanding dynamical systems.

DEFINITION 2.4.1. A metric space (X, d) is *disconnected* if there are two nonempty open sets U and V such that $X = U \cup V$ and $U \cap V = \emptyset$ (U and V are sometimes called a separation of X). If a metric space is not disconnected, it is *connected*.

Proving that a space is disconnected is usually much easier than showing it is connected. We can always build a disconnected space from two spaces X and Y as follows:

Let X and Y be metric spaces, and $Z = X \cup Y$. Fix two points $x_0 \in X$ and $y_0 \in Y$, and define the metric d_Z on Z by:

$$d_Z(z_1, z_2) = \begin{cases} d_X(z_1, z_2) & z_1, z_2 \in X \\ d_Y(z_1, z_2) & z_1, z_2 \in Y \\ d_X(z_1, x_0) + d_Y(z_2, y_0) + 1 & z_1 \in X, z_2 \in Y \\ d_Y(z_1, y_0) + d_X(z_2, x_0) + 1 & z_1 \in Y, z_2 \in X \end{cases}$$

We leave the following facts to the reader to check: d_Z is in fact a metric, and that a set is open if and only if it is a union of (possibly empty!) open sets from X and Y (this is a common feature of disconnected spaces). From here, it is easy to see that Z is disconnected, with the separation given by X and Y.

We give a few more examples:

EXAMPLE 2.4.1.

- $\mathbb{R} \setminus \{0\}$ is disconnected, since it is equal to $(-\infty, 0) \cup (0, \infty)$.
- Suppose that $x \in X$ is an isolated point. Writing $X = (X \setminus \{x\}) \cup \{x\}$ shows that X is disconnected if X has at least 2 points and an isolated point.
- If X is disconnected, then $X \times Y$ is disconnected for any metric space Y (the separation is given by $U \times Y$ and $V \times Y$, where U and V are the separating sets for X).

By contrast, we will spend some time showing that fundamental spaces like $[0, 1]$ and $\mathbb{R}$ *are* connected.

THEOREM 2.4.2. *Let X be a metric space. Then the following are equivalent:*

(i) X is connected
(ii) If $A \subset X$ is both open and closed, then either $A = \emptyset$ or $A = X$
(iii) If $\varphi : X \to \{0, 1\}$ is a continuous function, then φ is a constant function

PROOF. We will show that $(i) \implies (ii) \implies (iii) \implies (i)$. First, assume that X is connected, and that A is an open and closed subset of X. Then $X \setminus A$ is open because A is closed. So $X = A \cup (X \setminus A)$, and X cannot be connected.

Now suppose that X has no proper subsets which are simultaneously open and closed, and that $\varphi : X \to \{0, 1\}$ is continuous. Note that $\{0\} \subset \{0, 1\}$ is both open and closed in $\{0, 1\}$, so $\varphi^{-1}(0)$ is both open and closed in X by properties of continuity. Thus, either $\varphi^{-1}(0)$ is either X or $\emptyset$. If it is X, then φ is identically 0. If it is $\emptyset$, φ is identically 1.

Finally, assume that property (iii) holds. We will show connectedness. Assume, for a contradiction, that $X = U_1 \cup U_2$, with U_1 and U_2 open. Then define $\varphi : X \to \{0, 1\}$ by:

$$\varphi(x) = \begin{cases} 0 & x \in U_1 \\ 1 & x \in U_2 \end{cases}$$

One can easily check that φ is a continuous map, contradicting property (iii).

□

PROPOSITION 2.4.3. *A subset $U \subset \mathbb{R}$ is open if and only if U can be written as $U = \bigcup_{i=1}^{n} I_i$ (where $n \in \mathbb{N} \cup \{\infty\}$), where each I_i is an open interval or ray and $I_i \cap I_j = \emptyset$ when $i \neq j$.*

The strategy for proving this proposition will be to produce the intervals I_i by choosing a point inside them. We may imagine we place two tiny explorers at a point $x \in U$ and sending in each direction to walk until they find a point outside U. If points in U represent dry land and points outside U represent the ocean, our explorers mark the coasts of their "island." The coastlines are exactly the interval containing x! We make this precise in the following

Proof. Let $U \subset \mathbb{R}$ be a nonempty open set and $x \in U$. Then let $a_x = \inf\{a < x : (a, x) \subset \mathbb{R}\}$ and $b_x = \sup\{b > x : (x, b) \subset \mathbb{R}\}$ (the case when $a_x = -\infty$ or $b_x = \infty$ corresponds to a ray. Note that since U is open, there exists a $\varepsilon > 0$ such that $(x-\varepsilon, x+\varepsilon) \subset U$. Hence, $a_x < x < b_x$. We claim that $(a_x, b_x) \subset U$. We already know that $x \in U$. Choose any $z \in (a_x, b_x)$. Assume, without loss of generality, that $z \in (a_x, x)$ (the other case is completely symmetric). Then let $z' = (a_x + z)/2$, so that $z' > a_x$. Then $(z', x) \subset U$ by assumption. But $z \in (z', x)$, so $z \in U$.

We claim that if $y \in (a_x, b_x) \cap U$, then $a_y = a_x$ and $b_y = b_x$. Let us handle the case when $y < x$ (the opposite case is completely symmetric). We first show that $a_x = a_y$. Note that since $y \in (a_x, b_x)$ and $y < x < b_x$, $y \in (a_x, x)$. Note that $(a_x, y) \subset (a_x, x) \subset U$, so by definition of a_y, $a_y \leq a_x$. If $a_y < a_x$, then $(a_y, y) \subset U$. But since $a_x < y$, $(a_y, x) = (a_y, y) \cup (a_x, x)$, so $(a_y, x) \subset U$, contradicting the minimality of a_x.

Now we show that $b_x = b_y$. We continue to assume that $y < x$. Note that $b_y \geq b_x$, since if $(x, b) \subset U$, then $(y, b) = (y, x] \cup (x, b) \subset U$. Assume that $b_y > b_x$, to reach a contradiction. Then $(y, b_y) \subset U$. But $(x, b_y) \subset (y, b_y) \subset U$, contradicting the minimality of b_x. So $b_y = b_x$.

It is obvious that $U = \bigcup_{x \in U}(a_x, b_x)$. We have shown that if $(a_x, b_x) \cap (a_y, b_y) \neq \emptyset$, then $a_x = a_y$ and $b_x = b_y$. Thus, the union expressed is a disjoint union if we pick a representative for each interval or ray. We choose a rational number to represent each interval or ray. Since there are countably many rational numbers, there can be only countably many intervals and rays. Thus, U is a union of at most countably many disjoint open intervals and/or rays.

$\square$

Theorem 2.4.4. *$\mathbb{R}$ is connected*

Proof. We claim that if U_1 and U_2 are disjoint, nonempty subsets of $\mathbb{R}$, then $\mathbb{R} \neq U_1 \cup U_2$. Suppose otherwise. Indeed, neither U_1 nor U_2 can be all of $\mathbb{R}$. So U_1 must contain a proper interval or ray, which must have at least one endpoint in $\mathbb{R}$. Suppose that one of the intervals or rays making up U_1 can be written as (a, b), with $b \in \mathbb{R}$. Since $\mathbb{R} = U_1 \cup U_2$, either $b \in U_1$ or $b \in U_2$. In either case, we have that there exists a $\varepsilon > 0$ such that $B_\varepsilon(b) \subset U_i$. If it is in U_2, this contradicts disjointness of U_1 and U_2. If it is in U_1, then (a, b) is not one of the disjoint intervals making up U_1, since $(a, b + \varepsilon) \subset U_1$ (see the proof of Proposition 2.4.3).

$\square$

Corollary 2.4.5. *Intervals and Rays are connected.*

Proof. Suppose that X is an interval or ray, with endpoints a and b (possibly being $\pm\infty$). Suppose that $X = V_1 \cup V_2$, with V_1 and V_2 disjoint open subsets of X. We produce a separation of $\mathbb{R}$ by modifying the sets V_1 and V_2.

(a) **An endpoint is finite and open.**

Assume that a is an open right endpoint. Pick some $x_0 \in (a, b)$. Use an increasing function $f : (a, x_0] \to (-\infty, x_0]$, and extend it using the identity on (x_0, b) (or, if necessary, $(x_0, b]$). One could, for instance, take:

$$f(x) = x_0 + \frac{x - x_0}{x - a}$$

Then, we produce a separation of $(-\infty, b)$ (or $(-\infty, b]$)) by using $U_1 = f(V_1)$ and $U_2 = f(V_2)$ (U_1 and U_2 must be unions of disjoint open intervals since f is increasing).

(b) **An endpoint is finite and closed.**
If a is the closed left endpoint, then one of the open sets V_1 or V_2 must contain it. Assume without loss of generality that it is V_1. Then let $U_1 = (-\infty, a) \cup V_1$. Then U_1 and V_2 are open and separate $(-\infty, b)$ (or $(-\infty, b]$).

(c) **An endpoint is infinite (and hence open).**
In this case we do nothing.

Applying these modifications and the symmetric ones to the right endpoint allows us to build from a separation of the interval, a separation of $\mathbb{R}$, which is a contradiction to the connectedness of $\mathbb{R}$. So we must conclude that the interval or ray could not be separated.

□

THEOREM 2.4.6. *If $A \subset \mathbb{R}$ is a nonempty connected subset, then A is a point, interval, ray, or all of $\mathbb{R}$.*

PROOF. Let $a = \inf A$ and $b = \sup A$. If a or b is finite, we claim that without loss of generality, we may assume $a, b \in A$. Indeed, if $\varphi : A \cup \{a, b\} \to \{0, 1\}$ is continuous, then $\varphi|_A$ must be constant since A is connected. But then since $a = \inf A$, there exists a sequence $a_n \in A$ such that $a_n \to a$. Hence $\varphi(a) = \varphi(A)$. Similarly, $\varphi(b) = \varphi(A)$. Hence any function from $A \cup \{a, b\}$ must be constant. Since adding the infemum or supremum to and interval or ray makes it a closed interval or ray, we can assume that $a, b \in A$ (if they are finite).

If $a = b$, we are done, since in this case A is a point. Suppose that there exists some $x \in (a, b)$ such that $x \notin A$. Then $B = [a, x] \cap A = (a - 1, x) \cap A$, so B is both open and closed. Hence if A is connected, we conclude that A contains all points between a and b. That is, A is one of the prescribed sets. □

DEFINITION 2.4.7. Let X be a metric space. A *path connecting x and y in X* is a continuous function $f : [0, 1] \to X$ such that $f(0) = x$ and $f(1) = y$. X is *path-connected* if and only if for any two points $x, y \in X$, there is a path connecting x and y.

THEOREM 2.4.8. *Path connected spaces are connected.*

PROOF. Suppose that X is a space that is path connected, and assume for a contradiction that $X = U_1 \cup U_2$, with U_1 and U_2 open and nonempty. Choose $x \in U_1$ and $y \in U_2$. Since X is path connected, there exists a path $f : [0, 1] \to X$ such that $f(0) = x$ and $f(1) = y$. Let $V_1 = f^{-1}(U_1)$ and $V_2 = f^{-1}(U_2)$. Then since $U_1 \cup U_2 = X$, $V_1 \cup V_2 = [0, 1]$. Since U_1 and U_2 are disjoint, so are V_1 and V_2. Note that $0 \in V_1$ and $1 \in V_2$, so both are nonempty and open. Hence they give a separation of $[0, 1]$, a contradiction.

□

A reasonable guess one might make for metric spaces is that the converse is true. The following counterexample shows otherwise:

2.4.1. The Topologist's Sine Curve. Consider $g : (0, 1) \to [-1, 1]^2$ defined by $g(x) = (x, \sin(\pi/x))$. Note by Exercise 4 that the image of g is connected, and by Exercsie 6, the closure $X = \overline{g(0, 1)}$ is also connected. We leave in the exercises that X is not path connected, but it is intuitively clear when one sees that $\{0\} \times [-1, 1] \subset X$.

2.4.2. The Cantor Topology. We have seen two ways of detecting disconnected spaces: they may have an isolated point, or be the union of finitely many connected pieces. One might reasonably guess that this picture is true universally: that any space X is a union of pairwise disjoint open sets, each of which is connected. However, let us try to find a procedure to exhibit a counterexample.

Let $X_0 = [0, 1]$, the interval. We begin by chopping it in 2 pieces, but we remove a definite piece, not just a point. That is, $X_1 = X_0 \setminus (1/3, 2/3)$. Then $X_1 = [0, 1/3] \cup [2/3, 1]$ is still a union of finitely many pieces, so we divide it again, in a symmetric way: we remove the middle third of each

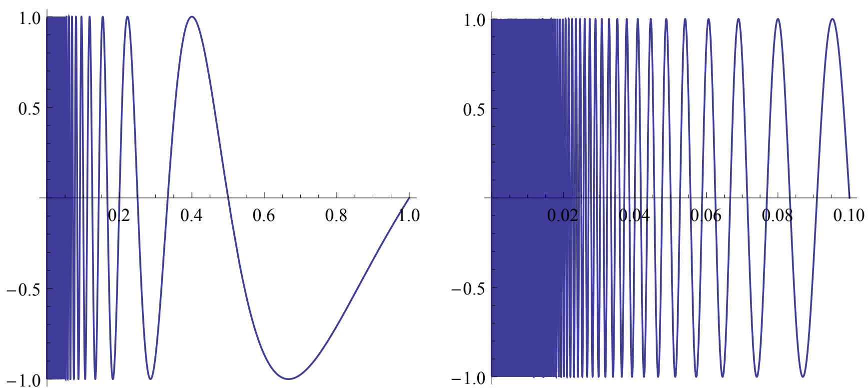

FIGURE 3. The Topolgist's Sine Curve plotted on $[0,1]$ and $[0,.1]$

interval, so that $X_2 = X_1 \setminus ((1/9, 2/9) \cup (7/9, 8/9))$. We see a way to iterate this process: X_n will consist of 2^n intervals, and to arrive at X_{n+1}, we remove the (open) middle third of each interval making up X_n. Then we let $X = \bigcap_{n=0}^{\infty} X_n$.

Let's make a few observations:

(i) **X is closed and nonempty**

Since X is an intersection of closed sets, it is closed. Furthermore, notice that if $x \in X_n$ is an *endpoint* of an interval, then $x \in X$, so, for instance $0 \in X$.

(ii) **X has no isolated points**

We need to show that for every $\varepsilon > 0$ and $x \in X$, there exists $y \in X$ such that $d(x, y) < \varepsilon$. Note that if $x \in X$, then $x \in X_n$ for every n. Choose n large enough so that $3^{-n} < \varepsilon$. Since X_n consists of a union of 2^n intervals of length 3^{-n}, x is in an interval of length 3^{-n}. By choosing one of its endpoints, we find a point $y \in X$ such that $d(x, y) < 3^{-n} < \varepsilon$.

(iii) **If $A \subset X$ is connected and nonempty, then $A = \{x_0\}$ for some x_0**

Suppose that A is connected, and $x \in A$. Assume, for a contradiction that $y \in A$ is a distinct point. Then $d(x, y) = \delta > 0$, so we can find an n such that $3^{-n} < \delta$. Then x and y lie in different intervals of the set X_n. If I is the interval containing x, note that if $U = B_{3^{-(n+1)}}(I) \subset [0, 1]$, then $U \cap X_n = I$. Then note that $U \cap A = I \cap A$, so this set is both open and closed in A. Hence, A is not connected.

Any space satisfying the last property is called *totally disconnected*. If X is a metric space which is compact, totally disconnected, and has no isolated points, then X is a called a *Cantor space*. Cantor spaces will arise naturally in our study of dynamical systems. In fact, it is true that every Cantor space is homeomorphic to the one constructed above, though we will not prove this fact.

Exercises.

1. Show that X is connected if and only if for any discrete space Y, and continuous function $f : X \to Y$, f is a constant function.
2. Show that if X and Y are connected spaces, then $X \times Y$ is connected.
3. Show that if $X = \bigcup_{i=1}^{n} X_i$, $X_i \cap X_j \neq \emptyset$ for every i, j and X_i is a connected subset of X for every i, then X is connected.
4. If $\mathcal{U}$ is an open cover of a metric space X, a *refinement* of $\mathcal{U}$ is an open cover $\mathcal{V}$ with the property that if $U \in \mathcal{U}$, there exists $V \in \mathcal{V}$ such that $V \subset U$. Show that any open cover of the Cantor set has a refinement by disjoint open sets.
5. Show that if X is connected and $f : X \to Y$ is continuous, then $f(X)$ is continuous.

6. Prove the Intermediate Value Theorem: If $f : [x, y] \to [a, b]$ is continuous, and $c \in [f(t_1), f(t_2)]$, then there exists $t \in [t_1, t_2]$ such that $f(t) = c$.
7. Show that if X is a metric space, and $A \subset X$ is a connected subset, then $\overline{A}$ is connected.
8. Show that if X is a metric space with countably many points, then X is totally disconnected. [*Hint*: For each $x \in X$, consider the function $d_x : X \to \mathbb{R}$ defined by $d_x(y) = d(x, y)$]
9. In this set of exercises, we show that the topologist's sine curve (denote it by X) is not path connected.
   (a) If x_k is a sequence in $[-1, 1]$, show that there exists a sequence $t_k \to 0$ such that $\sin(\pi/t_k) = x_k$.
   (b) Let $x = (1, 0) \in \mathbb{R}^2$ and $y = (0, 0)$. Show that $y \in X$.
   (c) Show that if $f : [0, 1] \to X$ is a path connecting x and y, and f_1 and f_2 are its first and second components, then if $f_1(s) > 0$, then $f_2(s) = \sin(\pi/f_1(s))$.
   (d) Use the Intermediate Value Theorem and the fact that $f_1(0) = 1$ and $f_1(1) = 0$ to construct a sequence s_k such that $f_1(s_k) = t_k$.
   (e) Show that there does not exist a path connecting x and y.
10. * Let $X = \bigcup_{i=1}^{n} X_i$ where each X_i is connected. Let A be the matrix such that $a_{ij} = 1$ if $X_i \cap X_j \neq \emptyset$ and $a_{ij} = 0$ if $X_i \cap X_j = \emptyset$. Let $a_{ij}^{(n)}$ denote the n^{th} entry of the matrix A^n. Show that X is connected if *and only if* for every pair i, j, there exists an $n \in \mathbb{N}$ such that $a_{ij}^{(n)} \neq 0$.

2.5. Topological Groups

Many of the topological spaces that we will study have additional structure. For instance, in $\mathbb{R}^k$, we have the addition operation, which is a continuous map. If we consider $S^1 \subset \mathbb{C} \cong \mathbb{R}^2$, then the product of two complex numbers is again a complex number of modulus one. That is, multiplication of two elements of S^1 is again an element of S^1, and the multipliaction map is continuous. We formalize this using the following:

Definition 2.5.1. A *group* is a pair (G, m), where G is a set and $m : G \times G \to G$ (we will often say $m(a, b) = ab$ or $m(a, b) = a \cdot b$) satisfies:

(a) $a \cdot (b \cdot c) = (a \cdot b) \cdot c$ for every $a, b, c \in G$
(b) There exists $e \in G$ such that $a \cdot e = a = e \cdot a$ for every $a \in g$
(c) For every $a \in G$, there exists a unique $b \in G$ (denoted a^{-1}) satisfying $ab = e$

If the group G also has a metric $d : G \times G \to \mathbb{R}$, it is a *topological group* if the map $m : G \times G \to G$ is continuous and the map $I : G \to G$ mapping $I(g) = g^{-1}$ is continuous.

Example 2.5.1.

1. $(\mathbb{R}, +)$ is a group and a topological group.
2. $(\mathbb{R}^n, +)$ is a group (under vector addition) and a topological group (with the usual Euclidean distance).
3. $(S^1, \cdot)$ is a group and a topological group.
4. $GL(n, \mathbb{R}) = \{A : A \text{ is a } n \times n \text{ matrix with } \det(A) \neq 0\}$ is a group. It is a topological group if we use the metric $d(A, B) = \sqrt{\sum_{1 \leq i,j \leq n} |a_{ij} - b_{ij}|^2}$.
5. Any group G is a topological group if it is given the discrete metric.

A group G is *abelian* if $ab = ba$ for every $a, b \in G$. In this case, we will usually use a few different notations. We write the group mulitplication using additive notation (ie, as $a + b$), and the group inverse using the additive inverse (ie, writing the inverse of a as $-a$). These notations come naturally from matrices, where addition is abelian but two matrices may not necessarily commute. If G is written with multiplication we will write g^n for the n-fold product of the element g, and if G is abelian, we will write it as ng.

Definition 2.5.2. If G is a group, a *subgroup* of G is a nonempty subset $H \subset G$ such that if $a, b \in H$, $a \cdot b^{-1} \in H$. H is *proper* if $H \neq \{e\}$ and $H \neq G$.

Example 2.5.2.

1. If G is any group, $\{e\} \subset G$ is called the *trivial* subgroup.
2. $\mathbb{Z}$ and $\mathbb{Q}$ are subgroups of $\mathbb{R}$
3. If G is an abelian group then $\mathbb{Z}g = g\mathbb{Z} = \{ng : n \in \mathbb{Z}\}$ is a subgroup (Exercise 1)
4. $SL(n, \mathbb{R}) = \{A \in GL(n, \mathbb{R}) : \det(A) = 1\}$ is a subgroup of $GL(n, \mathbb{R})$
5. Consider the group $(\mathbb{R} \setminus \{0\}, \times)$. Then $\mathbb{Q} \setminus \{0\}$ is a subgroup, but $\mathbb{Z} \setminus \{0\}$ is not.

A *closed subgroup* of a topological group G is a subgroup that is closed. A subgroup is *proper* if it is not the trivial subgroup or all of G.

Proposition 2.5.3. *If $H \subset G$ is a subgroup, then $\overline{H}$ is a closed subgroup*

Proof. Suppose that $a, b \in \overline{H}$. We want to show that $ab^{-1} \in \overline{H}$. Choose sequences $a_n \to a$ and $b_n \to b$ which lie completely in H. Such choices are possible by the definition of $\overline{H}$. But then $a_n b_n^{-1} \in H$ (since H is a subgroup), and $a_n b_n^{-1} \to ab^{-1}$ by continuity of group multiplication and inversion. Hence $ab^{-1} \in \overline{H}$. □

What can we say about a closed subgroup $H \subset \mathbb{R}$? We have already come up with a few examples: $H = \{0\}, \alpha\mathbb{Z}$, and $\mathbb{R}$. In fact, in general we have the following:

Theorem 2.5.4. *If H is a proper closed subgroup of G, there exists a unique $\alpha > 0$ such that $H = \alpha\mathbb{Z} = \{\alpha n : n \in \mathbb{Z}\}$*

The proof is best understood with a picture (Figure 4). We consider the set of all $\alpha > 0$ in H. If there were no smallest positive α, then we would have a sequence of points $a_n \to 0$ with $a_n > 0$ and $a_n \in H$. Then $\mathbb{Z}a_n \subset H$, because we can add or subtract a_n as often as we please. We can think of each subgroup as passing over $\mathbb{R}$ like a comb, and as the a_n get smaller, the comb's teeth get closer and closer together. In a sense, these subgroups (all of which are contained in H) "converge" to $\mathbb{R}$.

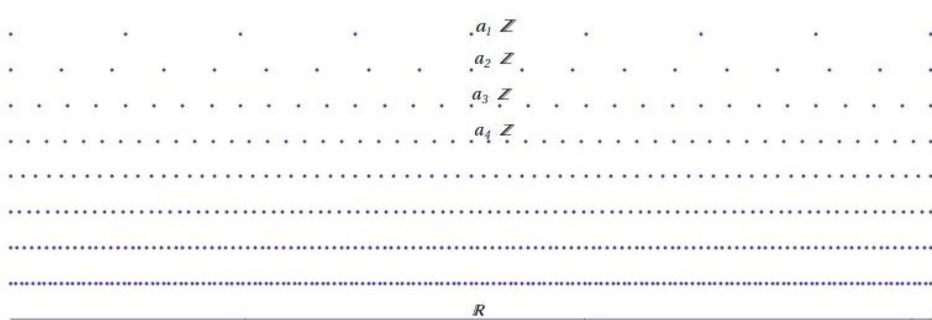

Figure 4. Subgroups of $\mathbb{R}$ "approaching" $\mathbb{R}$

Lemma 2.5.5. *If $\alpha > 0$ and $x \in \mathbb{R}$, then there exists an $n \in \mathbb{Z}$ such that $d(n\alpha, x) \leq \alpha/2$.*

Proof. Let $y = x/\alpha$. Then let n be the floor or ceiling of y, depending on which is closer. Then $|y - n| \leq 1/2$. Then $|x/\alpha - n| \leq 1/2$ and mutiplying both sides by α gives $|x - n\alpha| \leq \alpha/2$. □

Proof of Theorem 2.5.4. It is easy to see that if the α prescribed by the theorem exists, it must be unique. We know that H is not $\{0\}$ or $\mathbb{R}$, since it is proper. Considering H as a subset of $\mathbb{R}$, we can set $\alpha = \inf H \cap (0, \infty)$ (this set is nonempty by Exercise 3). We claim that $\alpha > 0$. We know since $H \cap (0, \infty) \subset (0, \infty)$, $\alpha \geq 0$. Assume, to obtain a contradiction, that $\alpha = 0$. Then there exists a sequence $a_n \to 0$ such that $a_n > 0$ and $a_n \in H$. We will show that if $x \in (0, \infty)$, then $x \in H$ (which contradicts $H \neq \mathbb{R}$).

For each a_n, let m_n be the integer produced by Lemma 2.5. Then $m_n a_n \in H$, since it can be written as $a_n + a_n + \cdots + a_n$, adding a_n to itself m_n times. Now, we know that $d(m_n a_n, x) \leq a_n/2 \to 0$, so x is the limit of a sequence of points in H. Since H is closed, $x \in H$, and $H = \mathbb{R}$.

Hence, $\alpha > 0$. We know that $\alpha\mathbb{Z} \subset H$, so we want to show that $H \subset \alpha\mathbb{Z}$. Choose $x \in H \cap (0, \infty)$. By Lemma , there exists an n such that $|x - n\alpha| \leq \alpha/2$. But then if $x - n\alpha > 0$, we have produced an element of H less than α, a contradiction. If $x - n\alpha < 0$, then $-(x - n\alpha) = |x - n\alpha|$ produces the same contradiction, since the subgroup H must be closed under additive inverse.

□

Let us see how this relates to the circle. Recall that $S^1 = \{z \in \mathbb{C} : |z| = 1\} \subset \mathbb{C} \setminus \{0\}$. Then S^1 is a closed subgroup of $(\mathbb{C} \setminus \{0\}, \cdot)$. We need a new notion to find an alternative characterization:

Definition 2.5.6. If G_1 and G_2 are groups, $\varphi : G_1 \to G_2$ is a *homomorphism* if $\varphi(ab) = \varphi(a)\varphi(b)$ for every $a, b \in G_1$. A bijective homomorphism is an *isomorphism*. Two groups are *isomorphic* if there is an isomorphism between them.

Some easy elementary properties of group homomorhpisms are that $\varphi(e_{G_1}) = e_{G_2}$ and $\varphi(a)^{-1} = \varphi(a^{-1})$. We leave these properties as exercises for the reader.

Example 2.5.3.

1. Define for $c \in \mathbb{R}$, $h_c : (\mathbb{R}, +) \to (\mathbb{R}, +)$ by $h_c(x) = cx$. Then by the distributive property, h_c is a homomorphism
2. Define $\exp : (\mathbb{R}, +) \to (\mathbb{R} \setminus \{0\}, \cdot)$ in the usual way. Then by elementary properties of the exponential, this is a homomorphism.
3. $\exp : (\mathbb{C}, +) \to (\mathbb{C} \setminus \{0\}, \cdot)$ is also a homomorphism.

Definition 2.5.7. Given a set X and an equivalence relation $\sim$, the *quotient space* $X/\sim$ is $X/\sim = \{[x] : x \in X\}$, where $[x]$ denotes the equivalence class of x. Each comes with a projection map $p : X \to X/\sim$ by sending $x \mapsto [x]$.

We use this notion to form a new characterization of the circle. Define an equivalence relation on $[0, 1]$ by setting $x \sim y$ if and only if $x = y$ or $x = 0$ or 1 and $y = 0$ or 1. That is, we identify 0 and 1 to be the same point. We can think of this as "gluing" 0 and 1. The justifies Exercise 12 of Section 1, and gives us good reason for thinking of the circle in this way!

Definition 2.5.8. Let G be a group and H be a subgroup. Define the equivalence relation on G by $a \sim b$ if and only if $a^{-1}b \in H$. We denote the quotient space $G/\sim$ by G/H. The equivalence class of an element of a is often written aH and is called a *coset* of H.

Proposition 2.5.9. *If G is an abelian group, then G/H is a group, with the group operation given by $[a] + [b] = [a + b]$. Furthermore, the map $p : G \to G/H$ given by $p(a) = [a]$ is a surjective homomorphism.*

Proof. We need to see first that the group operation is well-defined. Suppose that $a' \in [a]$ and $b' \in [b]$. We then need to show that $a' + b' \in [a + b]$. This is equivalent to showing that $-(a' + b') + (a + b) \in H$. But $-(a' + b') + (a + b) = (a - a') + (b - b')$, since G is abelian (and hence we can rerrange the order of addition). Now, since $a' \in [a]$, $a - a' \in H$. Similarly, $b - b' \in H$. Hence $(a + b) - (a' + b') \in H$.

We leave it to the reader to check that this new operation is in fact an abelian group operation (that is, that it is associative, has an identity, inverses, and is abelian). The map p is obviously surjective. It is a homomorphism because $p(a) + p(b) = [a] + [b] = [a + b] = p(a + b)$.

□

Consider the quotient group $\mathbb{R}/\mathbb{Z}$. We want to show that as a set, we can think of it as $[0, 1]/\{0 \sim 1\}$ (here, $\{0 \sim 1\}$ denotes the relation which "glues" only 0 and 1). Let $x \in \mathbb{R}$. Then there exists a unique $n \in \mathbb{Z}$ such that $x - n \in [0, 1)$. That is, $[x] \cap [0, 1) = (x + \mathbb{Z}) \cap [0, 1)$ consists of exactly one point. Thus, if we add the point 1 to $[0,1)$ to form $[0,1]$, the only nontrivial identification that must be made is $0 \sim 1$, as claimed.

Note that by Proposition 2.5.9, $\mathbb{R}/\mathbb{Z}$ is a group, with operation $[\theta]+[\rho]=[\theta+\rho]$. The following theorem will give us the final piece of the algebraic puzzle:

THEOREM 2.5.10 (First Isomorphism Theorem). *Let G and H be abelian groups, and $\varphi : G \to H$ be a sujective homomorphism. Then $\ker\varphi = \{g \in G : \varphi(g) = e\}$ is a subgroup, and H is isomorphic to $G/\ker\varphi$.*

PROOF. To see that $\ker\varphi$ is a subgroup, note that if $a, b \in \ker\varphi$, then $\varphi(a), \varphi(b) = e$. From the fact that φ is a homormophism, we know that $\varphi(ab^{-1}) = \varphi(a)\varphi(b)^{-1} = e$. Hence $ab^{-1} \in \ker\varphi$, and $\ker\varphi$ is a subgroup.

Let us construct the isomorphism Φ. For convenience, let $K = \ker\varphi$. Let $[a] \in G/K$, and define $\Phi([a]) = \varphi(a)$. If the map is well-defined it is obviously surjective. Let $b \in [a] \subset G$. Then $b^{-1}a \in K$, so $\varphi(b^{-1}a) = e$. Thus, $\varphi(b)^{-1}\varphi(a) = e$, and $\varphi(a) = \varphi(b)$. Thus Φ depends only on the equivalence class of a.

Furthermore, if $\Phi([a]) = \Phi([b])$, then $\varphi(a) = \varphi(b)$ and $\varphi(a^{-1}b) = e$, so $a \sim b$. That is, Φ is injective. $\square$

Φ is sometimes called the *map induced by φ*.

PROPOSITION 2.5.11. *The groups S^1 and $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ are isomorphic under the map $E : \mathbb{T} \to S^1$ defined by $E([t]) = \exp(2\pi it)$.*

PROOF. Observe that the map $\widetilde{E} : \mathbb{R} \to S^1$ defined by $\widetilde{E}(t) = \exp(2\pi it)$ is a surjective homomorphism from $\mathbb{R}$ to S^1 whose kernel is $\mathbb{Z}$. $\square$

Proposition 2.5.11 is the central tool for understand the algebraic properties of S^1. The construction is much more general and universal if we have similar topological structures associated with the groups. If G is a topological abelian group, a *translation invariant metric d* is a metric on G such that $d(g+h_1, g+h_2) = d(h_1, h_2)$ for every $g, h_1, h_2 \in G$. A map between metric spaces $f : X \to Y$ is said to be *open* if $f(U)$ is open in Y whenever U is open in X.

PROPOSITION 2.5.12. *If H is a closed subgroup of a topological abelian group G, and d_G is a translation-invariant metric on G, the function:*

$$d_{G/H}([g_1],[g_2]) = \inf\{d_G(g_1+h_1, g_2+h_2) : h_1, h_2 \in H\}$$

is a translation invariant metric on G/H, and that with this metric, the projection $p : G \to G/H$ is continuous and open.

This metric is called the *metric induced by d*.

PROOF. First, we show that it is a metric. It is clearly nonnegative, since it is the infemum over nonnegative numbers. Furthermore, it is symmetric by the symmetry of d_G. It is also clear that $d_{G/H}([g],[g]) = 0$. So we need to check that if $d_{G/H}([g_1],[g_2]) = 0$, then $[g_1] = [g_2]$ and the triangle inequality.

Suppose that $d_{G/H}([g_1],[g_2]) = 0$. Then $\inf\{d_G(g_1+h_1, g_2+h_2) : h_1, h_2 \in H\} = 0$. Hence there exist $h_{1,n}$ and $h_{2,n}$ such that $d_G(g_1+h_{1,n}, g_2+h_{2,n}) \to 0$. Since d_G is translation-invariant and G is abelian, we can rearrange terms to get that $d_G(h_{1,n}-h_{2,n}, g_2-g_1) \to 0$. But this implies that $h_{1,n}-h_{2,n}$ converges to g_2-g_1. Since $h_{1,n}-h_{2,n} \in H$ for every n, and H is closed, we conclude that their limit g_2-g_1 is also in H. But this exactly means that $[g_1] = [g_2]$.

To see the triangle inequality, observe that:

$$\begin{aligned}
d_{G/H}([g_1],[g_2]) &= \inf\{d_G(g_1+h_1, g_2+h_2) : h_1, h_2 \in H\} \\
&\leq \inf\{d_G(g_1+h_1, g_3+h_3) + d_G(g_3+h_3, g_2+h_2) : h_1, h_2, h_3 \in H\} \\
&\leq \inf\{d_G(g_1+h_1, g_3+h_3) : h_1, h_3 \in H\} + \inf\{d_G(g_3+h_3, g_2+h_2) : h_3, h_2 \in H\} \\
&= d_{G/H}([g_1],[g_3]) + d_{G/H}([g_3],[g_2])
\end{aligned}$$

To see that the projection is continuous, observe that $d_{G/H}([g_1],[g_2]) \leq d_G(g_1,g_2)$. To see that it is open, suppose that $U \subset G$ is open, and that $[x] \in p(U)$. Then $x \in U$, and there exists $\varepsilon > 0$ such that $B_\varepsilon(x) \subset U$. We claim that $B_\varepsilon([x]) \subset p(U)$. Indeed, if $d([y],[x]) < \varepsilon$, then there exists an $h \in H$ such that $d_G(x, y+h) < \varepsilon$. Hence $y + h \in B_\varepsilon(x) \subset U$, and hence $[y] = [y+h] \in p(U)$.

□

The following shows that the isomorphism of Proposition 2.5.11 is also a homeomorphism:

PROPOSITION 2.5.13. *Suppose that $f : G \to H$ is a surjective, continuous homomorphism of topological groups with translation-invariant metrics. Then if $K = \ker f$, the induced isomorphism $F : G/K \to H$ is continuous when G/K is given the induced metric. If f is open, F is a homeomorphism.*

PROOF. We know that F is an isomorphism, so it must be bijective. Let $p : G \to G/K$ denote the projection from G to G/K. We first show continuity of F. Suppose that U is open in H. We claim that

$$F^{-1}(U) = p(f^{-1}(U)). \tag{1}$$

Indeed, note that if $F([x]) \in U$, then $f(x) \in U$. So $F^{-1}(U) \subset p(f^{-1}(U))$. Furthermore, if $x \in f^{-1}(U)$, then $f(x) \in U$. Hence $F(p(x))F([x]) = f(x) \in U$. That is, $p(x) \in F^{-1}(U)$, and $p(f^{-1}(U)) \subset F^{-1}(U)$. Now, if $U \subset H$, then by continuity of f, $f^{-1}(U)$ is open in G. Then since p is open, $p(f^{-1}(U))$ is open in G/K. So by (1), we get that F is continuous.

Now assume that f is open. Similar arguments to those above show that $F(U) = f(p^{-1}(U))$. Then since p is continuous, $p^{-1}(U)$ is open. Since f is open, $f(p^{-1}(U))$ is open. So $(F^{-1})^{-1}(U) = F(U)$ is open, and F^{-1} is also continuous.

□

2.5.1. The p-adic numbers. We give one last example of a topological group which has very different properties from those we've seen. Fix a prime number $p \geq 2$. Define function $D_p : \mathbb{Z} \to \mathbb{R}_{\geq 0}$ by:

$$D_p(n) = \begin{cases} \max\left\{k \geq 0 : p^k \text{ divides } |n|\right\} & n \neq 0 \\ \infty & n = 0 \end{cases}$$

Then let $d_p(m,n) = p^{-D_p(m-n)}$. It is clear that d_p is symmetric, since $D_p(m-n) = D_p(n-m)$. It is also 0 if and only if $m = n$. So we need to verify the triangle inequality. Let $l, m, n \in \mathbb{Z}$. If p^k divides $l - m$ and p^k divides $m - n$, then p^k divides $l - n$. This implies that $D_p(l-n) \geq \min\{D_p(l-m), D_p(m-n)\}$. Because D_p appears as a negative exponent in d_p, we get that:

$$d_p(l,n) \leq \max\{d_p(l,m), d_p(m,n)\} \leq d_p(l,m) + d_p(m,n)$$

Finally, note that d_p is translation-invariant (we leave it to the reader to do this computation). We need one more Lemma before finding a genuinely new group. This generalizes the situation of $\mathbb{Q}$ and $\mathbb{R}$:

LEMMA 2.5.14. *If G is an abelian topological group with a translation-invariant metric d_G, then the completion $\widehat{G}$ is a topological abelian group with the group operation given by:*

$$g_1 + g_2 = \lim_{n\to\infty} g_{1,n} + g_{2,n}$$

and the metric given by:

$$d_{\widehat{G}}(g_1, g_2) = \lim_{n\to\infty} d_G(g_{1,n}, g_{2,n})$$

where $g_{i,n} \in G$ are chosen to converge to g_i for $i = 1, 2$.

PROOF. We first show that the group operation is a group operation and is well-defined. First, observe that $g_{i,n}$ must be Cauchy in n for each fixed i, so for every $\varepsilon > 0$ there exists an N such that if $m, n \geq N$, $d(g_{i,m}, g_{i,n}) < \varepsilon/2$ for $i = 1, 2$. But then:

$$\begin{aligned}
d_G(g_{1,m} + g_{2,m}, g_{1,n} + g_{2,n}) &= d_G(g_{1,m}, g_{1,n} + g_{2,n} - g_{2,m}) \\
&\leq d_G(g_{1,m}, g_{1,n}) + d(g_{1,n}, g_{1,n} + g_{2,n} - g_{2,m}) \\
&= d_G(g_{1,m}, g_{1,n}) + d(g_{2,m}, g_{2,n}) \\
&< \varepsilon/2 + \varepsilon/2 = \varepsilon
\end{aligned}$$

Hence $g_{1,n} + g_{2,n}$ is Cauchy and must converge in the completion.

We leave to the reader that addition in $\widehat{G}$ is well-defined, still associative, has an identity, and has inverses (one just takes associated sequences in G and uses these properties in the original group G).

□

The resulting group when we complete $\mathbb{Z}$ with the metric d_p is denoted by $\mathbb{Z}_p$ and called the *p-adic numbers*. It is another example of a compact group:

PROPOSITION 2.5.15. *$\mathbb{Z}_p$ is a Cantor space*

PROOF. We must show that $\mathbb{Z}_p$ is compact, has no isolated points, and is totally disconnected.

To see compactness, observe that if $x \in \mathbb{Z}_p$, $x = \lim_{k\to\infty} m_k$, where $m_k \in \mathbb{Z}$. Thus, if x_n is a sequence in $\mathbb{Z}_p$, we can form a sequence of sequences $m_{k,n}$, where $\lim_{k\to\infty} m_{k,n} = x_n$. If $m_{n,n}$ has a convergent subsequence, m_{n_k,n_k}, then x_{n_k} will converge as well. Hence it suffices to show that any sequence $x_n \in \mathbb{Z}$ has a convergent subsequence in $\mathbb{Z}_p$.

Recall that any $x_n \in \mathbb{Z}$ can be written uniquely as $x_n = \pm \sum_{i=0}^{\infty} a_{i,n} p^n$ for some $a_{i,n} = 0, \dots, p-1$, and $a_{i,n} = 0$ for all $i \geq N$ (where N may depend on n). This is exactly the expansion in base p. Since there are infinitely many elements in the sequence, there must be either infinitely many positive or negative ones. Assume without loss of generality that all elements of the sequence are positive. Now, since there are only finitely many choices for $a_{0,n}$, one number in $\{0, \dots, p-1\}$ must appear infinitely often. Let a_0 be that number. Then consider all sequences that begin with the symbol a_0. Since this number occurs infinitely often, and there are only finitely many choices for the digit $a_{1,n}$ there must exist some a_1 that occurs infinitely often. Iterate this process to get a sequence of digits a_k, and elements x_{n_k} whose first k digits in the base p expansion match $y_k = \sum a_k p^k$. Observe then that p^{k+1} divides $x_{n_k} - y_k$, so $d(x_{n_k}, y_k) < p^{-(k+1)}$. Furthermore, y_k is Cauchy since $d(y_k, y_l) = p^{-\min\{k,l\}}$ (Check this!). Hence y_k, and likewise x_{n_k} converges to some $y \in \mathbb{Z}_p$.

To see that $\mathbb{Z}_p$ has no isolated points, it suffices to show that $\mathbb{Z}$ has no isolated points with the metric d_p. Indeed, for any integer n, $n + p^k \to n$ as $k \to \infty$.

Now we see that $\mathbb{Z}_p$ is totally disconnected. Suppose that $A \subset \mathbb{Z}_p$ is connected and nonempty, and let $x \in A$. Then consider the function $d_x : A \to \mathbb{R}$ defined by $d_x(y) = d_p(x, y)$. Then since A is connected, $d_x(A)$ is also connected, and hence either a point, an interval, a ray, or the entire line. But notice that the distance between any two points of $\mathbb{Z}_p$ must be a limit point of distances of points from $\mathbb{Z}$ with the metric d_p. But the metric d_p can only take values in $\{0\} \cup \{p^{-n} : n \geq 0\}$. This set is closed, hence $d_x(A) \subset \{0\} \cup \{p^{-n} : n \geq 0\}$. We know that 0 is in the image since $d_x(x) = 0$. Now we see that the image cannot be an interval, ray or the entire line, so we conclude that $d_x(A) = \{0\}$. But then for any $y \in A$, $d(x, y) = 0$ and therefore $y = x$. That is, $A = \{x\}$.

□

Exercises.

1. Let $(G, +)$ be an abelian group. Show that $\mathbb{Z}g$ is a subgroup.
2. Let $(G, \cdot)$ be a group. Show that $H \subset G$ is a subgroup if and only if $(H, \cdot|_{H\times H})$ is a group.
3. Show that if H_1 and H_2 are subgroups of $\mathbb{R}$, $H_1 = H_2$ if and only if $H_1 \cap (0, \infty) = H_2 \cap (0, \infty)$

4. Calculate the translation-invariant metric for a quotient group (Proposition 2.5.12) in the case when $G = \mathbb{R}$ has the usual metric, $H = \mathbb{Z}$, and we think of $\mathbb{R}/\mathbb{Z}$ as the set of points $[0, 1)$. Does this remind you of anything?
5. Find a surjective continuous homomorphism $f : G \to H$ of groups with translation-invariant metrics such that the induced homomorphism $F : G/\ker f \to H$ is an isomorphism, but not a homeomorphism.
6. Suppose that $f : G_1 \to G_2$ is a continuous homomorhpism of abelian groups. Show that:
   (a) If $H_1 \subset G_1$ and $H_2 \subset G_2$, then $\bar{f} : G_1/H_1 \to G_2/H_2$ defined by $\bar{f}([g]) = [f(g)]$ is well-defined if and only if $f(H_1) \subset H_2$.
   (b) If f is an isomorphism, show that $\bar{f}$ is an isomorphism if and only if $f(H_1) = H_2$.
   (c) If $\bar{f}$ is an isomorphism, show that $\bar{f}$ is a homeomorphism if and only if f is a homeomorphism.
7. Prove or find a counterexample: If G is a discrete abelian topological group, and $p : G \to H$ is a continuous homomorphism, then $p(G)$ is a discrete subgroup of H.
8. Prove or find a counterexample: The only dense proper subgroups of $\mathbb{R}$ are isomorphic $\mathbb{Q}$.
9. Show that $p\mathbb{Z}_p = \{pn : n \in \mathbb{Z}_p\} = \overline{\{pn : n \in \mathbb{Z}\}}$, $p\mathbb{Z}_p$ is a proper, closed subgroup of $\mathbb{Z}_p$, and $\mathbb{Z}_p/p\mathbb{Z}_p \cong \mathbb{Z}/p\mathbb{Z}$.
10. Find an explicit separation of $\mathbb{Z}_2$.
11. Construct a homeomorphism between $\mathbb{Z}_2$ and the Cantor set described at the end of Section 2.4.
12. Suppose that G is a countable group with a translation-invariant metric d_G. Show that if G is not discrete, then it is not complete.

CHAPTER 3

What is a Dynamical System?

This chapter moves our focus from topology to dynamical systems. Chapter **??** introduced fundamental notions we will use to describe static properties of spaces: distances, limits and relationships between spaces (via continuous maps and homorphisms). But continuous maps have another interpretation: that of an evolution law. If the target space of a continuous map coincides with its domain, we may consider an n-fold composition of that map, and interpret $f^n(x)$ as the position of x after n time intervals have passed. This interpretation is at the foundation of dynamical systems, and understanding the asymptotic behaviour and classification of these evolution laws is at the heart of the study of modern dynamical systems. In this chapter, we exhibit some examples and introduce certain invariants and properties of dynamical systems that will aid us in the study of general systems.

3.1. Definitions and First Examples

In this section, we introduce the notion of a dynamical system, and certain other elementary concepts.

DEFINITION 3.1.1.

(a) A *dynamical system* on a metric space X is a continuous map $T : X \to X$

(b) An *invertible dynamical system* on a metric space X is a homeomorphism $T : X \to X$

(c) For each $k \in \mathbb{Z}$, define the k^{th} iterate of T to be the function:

$$T^k = \begin{cases} \overbrace{T \circ T \circ \ldots \circ T}^{k \text{ times}} & k > 0 \\ \text{id} & k = 0 \\ \underbrace{T^{-1} \circ T^{-1} \circ \ldots \circ T^{-1}}_{-k \text{ times}} & k < 0 \end{cases}$$

In the noninvertible case, only the positive iterates are defined.

(d) If $x \in X$, the *orbit* of X is the bi-directional sequence $\mathcal{O}(x) = \{T^k(x) : k \in \mathbb{Z}\}$. The *forward orbit* or *positive orbit* is the sequence $\mathcal{O}^+(x) = \{T^k(x) : k \in \mathbb{N}_0\}$. The *backward orbit* or *negative orbit* is the sequence $\mathcal{O}^-(x) = \{T^k(x) : k \in -\mathbb{N}\}$. In the noninvertible case, we set $\mathcal{O}(x) = \mathcal{O}^+(x)$, and the backward orbit is not defined.

(e) A point x is called *periodic* if there exists $n \in \mathbb{Z}$ such that $T^n(x) = x$. In this case, the *period* of x is $p(x) = \min\{n \in \mathbb{N} : T^n(x) = x\}$. The set of periodic points will be denoted $\text{Per}(T)$.

(f) A point x is called *eventually periodic* for $T : X \to X$ if there exist $m \neq n \in \mathbb{N}$ such that $T^m(x) = T^n(x)$

(g) A point x is *fixed* if $T(x) = x$. The set of fixed points will be denoted $\text{Fix}(T)$

(h) A subset $\Lambda \subset X$ is *invariant* if $f^{-1}(\Lambda) = \Lambda$. Λ is called *forward-invariant* if $f(\Lambda) \subset \Lambda$.

In most cases we consider, X will be compact. We can already see the connection between dynamical system and metric topology by observing that if we fix a homeomorphism $h : X \to Y$, let $\mathcal{H}(X, Y)$ be the set of homeomorphisms from X to Y and $\text{Dyn}(X)$ the space of invertible dynamical systems on X, there is a bijection $\Phi_h : \mathcal{H}(X, Y) \to \text{Dyn}(X)$ by mapping $\Phi_h(k) = h^{-1} \circ k$. What is a formula for Φ_h^{-1}?

Example 3.1.1. Let $X = [0, 1]$ and $f : X \to X$ be given by $f(x) = 4x(1-x)$. This is a noninvertible dynamical system, and exhibits many types of orbits. We begin by calculating the iterates of f:

$$\begin{aligned}
f^0(x) &= x \\
f^1(x) &= 4x(1-x) \\
f^2(x) &= 4f(x)(1-f(x)) \\
&= 4 \cdot 4x(1-x)(1-4x(1-x)) \\
&= -64x^4 + 128x^3 - 80x^2 + 16x \\
f^3(x) &= 4f^2(x)(1-f^2(x)) \\
&= -16384x^8 + 65536x^7 - 106496x^6 + 90112x^5 - 42240x^4 + 10752x^3 - 1344x^2 + 64x
\end{aligned}$$

The graphs of the iterates are plotted below. Note that as we increase the iterate, the more complex the graph becomes. This is a common feature of systems on the interval with large slope.

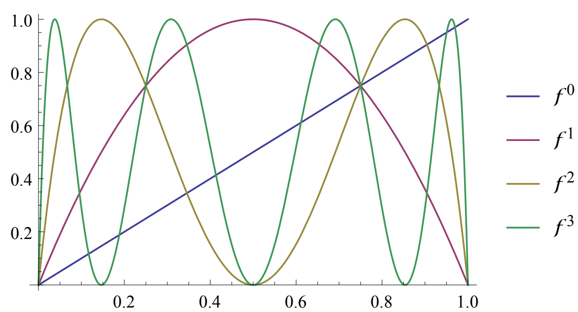

Figure 1. The first three iterates of f

To find the fixed points of f, we need to solve the equation $f(x) = x$, which as two solutions: $x = 0, 3/4$. The points of period 2 are exactly the points such that $f^2(x) = x$. We know that $x = 0$ and $x = 3/4$ will solve this equation (since, for instance, $f^2(0) = f(f(0)) = f(0) = 0$), so after factoring these solutions out, the equation becomes:

$$64x(x-3/4)\left(x - \frac{1}{8}(5-\sqrt{5})\right)\left(x - \frac{1}{8}(5+\sqrt{5}\right)$$

So $\left\{\frac{1}{8}(5 \pm \sqrt{5})\right\}$ are the points of period 2.

The peroidic points of f at any period must be finite, since $f^n(x)$ is always a polynomial of higher degree greater than 2, hence $f^n(x) - x$ always has a number of solutions less than or equal to the degree of f^n. This means that the number of periodic points is countable, since it is a countable union of finite sets. Since $[0, 1]$ contains uncountably many points, f must have an orbit which is not periodic.

Definition 3.1.2. Let $T : X \to X$ be an invertible dynamical system.

(a) If $x \in X$, the ω-limit set of x is the set $\omega(x) = \bigcap_{n\in\mathbb{N}} \overline{T^n(\mathcal{O}^+(x))}$. This is immediately extendable to the noninvertible case.
(b) If $x \in X$, the α-limit set of x is the set $\alpha(x) = \bigcap_{n\in\mathbb{N}} \overline{T^{-n}(\mathcal{O}^-(x))}$.

We observe that by Exercise 8 of Section 1, Chapter 1, $\omega(x)$ is exactly the set of limits of convergent subsequences of $\mathcal{O}^+(x)$ and $\alpha(x)$ is the set of limits of convergent subsequences of $\mathcal{O}^-(x)$. This allows us to make sense of the noninvertible case, by setting

$$\alpha(x) = \{y \in X : \exists n_1 < n_2 < \dots \text{ and } y_k \in T^{-n_k}(\{x\}) \text{ such that } y_n \to y\}$$

PROPOSITION 3.1.3. *Let $f : X \to X$ be an invertible dynamical system. Then for every $x \in X$, $\omega(x)$ and $\alpha(x)$ are closed, invariant subsets of X*

PROOF. We first show invariance. Let $m \in \mathbb{Z}$. Note that if $f^{n_k}(x) \to y$, then $f^m(f^{n_k}(x)) \to f^m y$, so $f^m(y) \in \omega(x)$. Thus $f^m(\omega(x)) = \omega(x)$.

To see that it is closed, let $w_n \in \omega(x)$ be a sequence converging to some w. Then there exist sequences $x_k^n \to w_n$ (here the sequence is indexed by k), where each x_k^n is some forward iterate of x which increase and become arbitrarily large. Without loss of generality, we can assume that $d(x_k^n, w_n) < 2^{-k}$. Let $y_n = x_n^n$. We claim $y_n \to w$. Let $\varepsilon > 0$. Choose N_1 such that $n \geq N_1$ implies $d(w_n, w) < \varepsilon/2$. Choose N_2 such that $2^{-N_2} < \varepsilon/2$. Then if $n \geq \max\{N_1, N_2\}$:

$$d(y_n, w) \leq d(x_n^n, w_n) + d(w_n, w) < \varepsilon/2 + \varepsilon/2 = \varepsilon$$

□

The following class of examples is crucial in the analysis of dynamical systems on the circle:

EXAMPLE 3.1.2. Let $R_\alpha : S^1 \to S^1$ be rotation by the angle $2\pi\alpha$ in the counter-clockwise direction. That is, if $z \in S^1$, then $R_\alpha(z)$ is the point on S^1 which in the counter-clockwise direction, makes an angle of $2\pi\alpha$ with z.

PROPOSITION 3.1.4. *R_α has two types of behavior:*

(i) If $\alpha = 2\pi r$ with $r \in \mathbb{Q}$, then every orbit of R_α is periodic of the same period
(ii) If $\alpha \notin 2\pi\mathbb{Q}$, then for every $x \in S^1$, $S^1 = \overline{\mathcal{O}(x)}$

It is our first example of a classification for a certain type of dynamical system (Circle Rotations). We will prove Proposition 3.1.4 in Section 2.

Exercises.

1. Show that if T is invertible, x is periodic if and only if x is eventually periodic.
2. Show that if Λ is invariant, it is forward-invariant. Show that if T is invertible, if Λ is forward-invariant, it is invariant.
3. Let T be invertible and $\mathrm{Per}_n(T)$ denote the points of period n. Show that $\mathrm{Per}_n(T) \subset \mathrm{Fix}(T^n)$, and that if n is prime, then $\mathrm{Per}_n(T) = \mathrm{Fix}(T^n) \setminus \mathrm{Fix}(T)$. Find a formula for $\mathrm{Per}_n(T)$ when n is not prime.
4. Assume that T is invertible and suppose that $\mathcal{O}(x)$ is a compact set. Show that x is periodic.
5. Let T be invertible and $x \in X$. Show that if $\omega(x)$ can be written as $\omega(x) = C_1 \cup C_2$ such that C_1 and C_2 are disjoint, closed, invariant sets, then either $C_1 = \emptyset$ or $C_2 = \emptyset$.
6. Let $T : [0,1] \to [0,1]$ be given by $T(x) = 2x$ if $x \in [0, 1/2]$ and $T(x) = 2 - 2x$ for $x \in [1/2, 1]$.
   (a) Draw the graph of T.
   (b) Show that for every $n \geq 1$, and for every $0 \leq k \leq 2^n - 1$, $k \in \mathbb{Z}$,
   $$T^n\left[\frac{k}{2^n}, \frac{k+1}{2^n}\right] = [0,1].$$
   (c) Show that the set of points which are eventually fixed by T are dense in $[0,1]$.
   (d) Show that $\mathrm{Per}(T)$ is dense in $[0,1]$ (*Hint*: Use the Intermediate Value Theorem)
   (d)* Show that there is a point $x \in [0,1]$ such that $\omega(x) = [0,1]$.
7. Let $T : [0,1] \to [0,1]$ be given by $T(x) = x^2$.
   (a) Show that $\mathrm{Per}(T) = \mathrm{Fix}(T) = \{0,1\}$.
   (b) Show that if $x \in (0,1)$, $\alpha(x) = \{1\}$ and $\omega(x) = \{0\}$.
   (c) Show that if $T : [0,1] \to [0,1]$ is an invertible dynamical system, then T is strictly monotone
   (d) Show that if $T : [0,1] \to [0,1]$ is an invertible dynamical system, then $\mathrm{Fix}(T) = \mathrm{Per}(T)$

(e) Show that if $T : [0,1] \to [0,1]$ is an invertible dynamical system with T increasing and $\mathrm{Fix}(T) = \{0,1\}$, then either $T(x) > x$ for every $x \in [0,1]$ or $T(x) < x$ for every $x \in [0,1]$.
(f) Show that if $T : [0,1] \to [0,1]$ is an invertible dynamical system such that:
 (i) T is increasing
 (ii) $T(x) < x$ for all $x \in [0,1]$
 (iii) $\mathrm{Fix}(T) = \{0,1\}$
 then $\alpha(x) = \{1\}$ and $\omega(x) = \{0\}$.

8. Show that if $x \in X$, $\omega(x) \cap \mathcal{O}^+(x) \neq \emptyset$ if and only if x is eventually periodic, and that $\omega(x)$ is the accumulation points of x if and only if x is not eventually periodic.

3.2. Transitivity and Minimality

Let us return to the case of the circle, $\mathbb{R}/\mathbb{Z}$ and the rotation T_α. By Proposition 3.1.4, we know that rotations on the circle have a rather simple classification: if one $x \in X$ has a dense orbit, then every $x \in X$ has a dense orbit.

Does this hold in general? The answer is no. For example, take the map $f : [0,1] \to [0,1]$ defined by:

$$f(x) = \begin{cases} 2x & x < 1/2 \\ 2 - 2x & x \geq 1/2 \end{cases}$$

What can we say about the set $\{x \in [0,1] : \omega(x) = [0,1]\}$? We know this dynamical system has many periodic orbits, so it cannot be all of $[0,1]$. However, it does form a dense-G_δ set:

Proposition 3.2.1. *Let X be a compact metric space and $T : X \to X$ a continuous map. The following are equivalent:*

(1) There exists some $x \in X$ such that $\omega(x) = X$
(2) For every pair of open sets U and V in X, there is some $n > 0$ such that $T^{-n}(U) \cap V \neq \emptyset$
(3) $A = \{x \in X : \omega(x) = X\}$ is a dense-G_δ set.

Any T satisfying one (and hence all) of these properties is called *topologically transitive*. To prove this this, we need one more fundamental notion:

Definition 3.2.2. A metric space X is *seperable* if there is a countable, dense subset of X

Lemma 3.2.3. *Every compact metric space is seperable.*

Proof. Let $\mathcal{U}_n = \{B_{1/n}(x)\}$ be the open cover of X by balls of radius $1/n$. Then since X is compact, there exists a finite subcover $X = B_{1/n}(x_1^n) \cup \cdots \cup B_{1/n}(x_{m_n}^n)$, where $x_1^n, \ldots, x_{m_n}^n$ is a finite set of points. Then let $A = \left\{x_i^j : j \in \mathbb{N}, i \in [1, m_j]\right\}$. Now choose $y \in X$, and let $y_n = x_i^n \in B_{1/n}(y)$ (such points exist by construction). Thus $y_n \to y$ and y a limit point of points in A. That is, $\overline{A} = X$. □

We are now ready to prove the equivalence of the conditions:

Proof of Proposition 3.3.2. Fix a countable dense set $D \subset X$. Consider the collection of balls $B_{k,n} = B_{1/k}(x_n)$ such that x_n is an enumeration of D. Since $\mathbb{N} \times \mathbb{N}$ is countable, we can, for convenience, write simply $B_{k,n} = B_m$. Note that a set L is dense in X if and only if for every m, $L \cap \overline{B_m} \neq \emptyset$.

Let us investigate the structure of $A = \{x \in X : \omega(x) = X\}$. Given B_m, let $W_{m,N} = \cup_{n \geq N} f^{-n}(B_m)$. Then $W_{m,N}$ is an open set. Furthermore, if $x \in W_{m,N}$, then there exists some $n \geq N$ such that $T^n(x) \in B_m$. Then let $R_m = \bigcap_{N \geq 0} W_{i,N}$. Note that R_m is a G_δ set. Furthermore, $x \in R_m$ if and only if for every N, there exists some $n \geq N$ such that $T^n(x) \in B_m$. That is, R_m is exactly the set $R_m = \{x \in X : \exists n_k \to \infty \text{ such that } f^{n_k}(x) \in B_m\} = \left\{x \in X : \overline{B_m} \cap \omega(x) \neq \emptyset\right\}$.

Since $\omega(x)$ is closed, it is dense in X if and only if $\omega(x) = X$. Furthermore, it is dense if and only if $\omega(x) \cap \overline{B_m} \neq \emptyset$ for every m. By the above arguments, this holds if and only if, $x \in \cap_{m \geq 1} R_m = A$. This is a G_δ set. The cycle of arguments will now follow swiftly:

1 $\implies$ **3**. Suppose that x_0 has $\omega(x_0) = X$. Then for any $n \geq 0$, $f^n(x_0) \in A$. Then $\{f^n(x_0)\} \subset A$ is dense. Thus A is dense, and we observed above that A is G_δ.

3 $\implies$ **1**. If a set is dense-G_δ it is nonempty.

2 $\iff$ **1**. See Katok-Hasselblatt, Lemma 1.4.2. □

Definition 3.2.4. Let X be a metric space. A point $x \in X$ is *isolated* if there is a $\varepsilon > 0$ such that $B_\varepsilon(x) \cap X = \{x\}$. X is said to be *perfect* if it has no isolated points.

Proposition 3.2.5. *Suppose $T : X \to X$ is a dynamical system on a perfect compact metric space X such that $\overline{\mathcal{O}(x)} = X$ for some x. Then T is topologically transitive.*

Proof. Since X is perfect, we only need to verify the $\mathcal{O}^+(x)$ is dense (which will imply that $\omega(x) = X$). Choose a sequence of integers n_k such that $|n_k|$ is increasing and $T^{n_k}(x) \to x$. Note that there exist infinitely many indices of n_k which are positive or infinitely many that are negative (or both). If infinitely many are positive, then since $|n_k| \to \infty$, and $T^{n_k+m}(x) \to T^m(x)$, we know that $\overline{\mathcal{O}(x)} \subset \overline{\mathcal{O}^+(x)}$. But since the full orbit is dense, $\overline{\mathcal{O}^+(x)} = X$.

Now consider the case when infinitely many indices are negative. Instead of $\overline{\mathcal{O}(x)} \subset \overline{\mathcal{O}^+(x)}$, we have $\overline{\mathcal{O}(x)} \subset \overline{\mathcal{O}^-(x)}$. Let U and V be open subsets of X. Since the full orbit is dense, the backward orbit is dense, and hence there exist $m < n < 0$ such that $T^m(x) \in U$ and $T^n(x) \in V$. Then $T^m(x) \in T^{m-n}(V) \cap U$, and T is topologically transitive by Proposition 3.3.2. □

We can't drop the perfectness condition. Let $X = \{a_n\} \cup \{0, 1\}$ such that a_n is an increasing sequence of points in X converging to 1 in the forward direction, and a_{-n} converges to 0. Then define $f(a_n) = a_{n+1}$. Here, the orbit of every point is dense in X, but the ω-limit set of every point is just $\{1\}$.

3.3. Semiconjugacy, Conjugacy, and Skew Products

To better understand the maps R_α, we will introduce new coordinates. But to make precise the notion of "new coordinates" in metric spaces, we need a new dynamical idea. We will see that this idea allows us to consider rotations by the angle of $2\pi\alpha$ through the maps $T_\alpha : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ defined by $T_\alpha(t) = t + \alpha$.

Definition 3.3.1. A *(topological) semiconjugacy* between dynamical systems $T : X \to X$ and $S : Y \to Y$ is a continuous map $f : X \to Y$ such that $f \circ T = S \circ f$.

$$\begin{array}{ccc} X & \xrightarrow{T} & X \\ f\big\downarrow & & \big\downarrow f \\ Y & \xrightarrow{S} & Y \end{array}$$

If f is a homeomorphism, then it is called a *(topological) conjugacy*. f is called the *conjugation* or *change of coordinates*. If f is continuous but not invertible, we say that S is a *factor* of T.

Topological conjugacy, like homeomorphisms for metric spaces, provides a natural equivalence relation on dynamical systems. We can motivate the terminology by noting that since f is a homeomorphism, T is conjugate to S if and only if $T = f^{-1} \circ S \circ f$ (think about matrix conjugacy). Doing some diagram chasing, we can see that this amounts to starting from a point $y \in Y$ (in the bottom left), we can go "upstairs" to a point of X (top left), apply T (top right) and then go back "downstairs" to Y (bottom right) to end up with the same result as just applying S. In this sense, T and S can be thought of as the same dynamical system.

We can now formally show that the maps T_α and R_α are conjugate:

PROPOSITION 3.3.2. *The map $\widetilde{E} : \mathbb{R} \to S^1$ defined by $\widetilde{E}(t) = \exp(2\pi i t)$ is a semiconjugacy between the dynamical systems $T_\alpha : \mathbb{R} \to \mathbb{R}$ defined by $T_\alpha(x) = x + \alpha$, and rotation by $2\pi\alpha$, $R_\alpha : S^1 \to S^1$. The induced map $E : \mathbb{R}/\mathbb{Z} \to S^1$ is a conjugacy.*

We need another notation and one more Proposition to prove Proposition 3.1.4. In general, if A and B are subsets of an abelian group, we set $A + B = \{a + b : a \in A, b \in B\}$.

PROPOSITION 3.3.3. *Let $\alpha \in \mathbb{R}$. Then $\alpha \in \mathbb{Q}$ if and only if $\mathbb{Z} + \alpha\mathbb{Z} = \frac{1}{q}\mathbb{Z}$, where $\alpha = \frac{p}{q}$ is in reduced form, and $\alpha \notin \mathbb{Q}$ if and only if $\overline{\mathbb{Z} + \alpha\mathbb{Z}} = \mathbb{R}$.*

PROOF. Assume that $\overline{\mathbb{Z} + \alpha\mathbb{Z}} \neq \mathbb{R}$. Then there exists some β such that $\mathbb{Z} + \alpha\mathbb{Z} = \beta\mathbb{Z}$ by . Now $1 = 1 + 0\alpha \in \mathbb{Z} + \alpha\mathbb{Z}$, so $1 = n\beta$ for some $n \in \mathbb{Z}$. But then $\beta = \frac{1}{n}$. But now $\alpha = 0 + 1\alpha \in \mathbb{Z} + \alpha\mathbb{Z}$, so $\alpha = m\beta$, and $\alpha = \frac{m}{n}$ for some $m, n \in \mathbb{Z}$.

So we need only to prove that if $\mathbb{Z} + \alpha\mathbb{Z} = \frac{1}{q}\mathbb{Z}$, then $\alpha = \frac{p}{q}$. Suppose that $\alpha = \frac{p}{q}$. It is obvious that $\mathbb{Z} + \frac{p}{q} \subset \frac{1}{q}\mathbb{Z}$, so we need only to show that $\frac{1}{q} \in \mathbb{Z} + \frac{p}{q}\mathbb{Z}$.

This amounts to finding integers $m, n \in Z$ such that $\frac{1}{q} = k + m\frac{p}{q}$, which is equivalent to $1 = kq + mp$. But since $\frac{p}{q}$ is in reduced form, p and q are coprime, and hence there exists a solution. □

PROOF OF PROPOSITION 3.1.4. We leave the case when $\alpha \in \mathbb{Q}$ as an exercise. Recall the projection $p : \mathbb{R} \to \mathbb{R}/\mathbb{Z}$, and consider:

$$p^{-1}(\mathcal{O}(0)) = p^{-1}(\alpha\mathbb{Z}) = \{k + n\alpha : k \in \mathbb{Z}, n \in \mathbb{Z}\} = \mathbb{Z} + \alpha\mathbb{Z}$$

We first observe that $\mathbb{Z} + \alpha\mathbb{Z}$ is a subgroup of $\mathbb{R}$. Now apply Proposition **??** to get that $p^{-1}(\mathcal{O}([x]))$ is dense if $\alpha \notin \mathbb{Q}$.

Let $[x], [y] \in \mathbb{R}/\mathbb{Z}$. We want to find a subsequence of of $\mathcal{O}([x])$ which converges to $[y]$. We can find a sequence $n_i + k_i\alpha \to y - x$ in $\mathbb{R}$, which means that $[k_i\alpha] \to [y - x]$, and $[x + k_i\alpha] \to [y]$. That is, there is a sequence of points in $\mathcal{O}([x])$ which converges to $[y]$.

□

3.3.1. Skew Products.

We define here a special type of factor which takes a particularly useful form. Let X and Y be metric spaces, and choose a continuous function $S : X \to C(Y, Y)$. We will denote $S(x)$ by S_x, so that $S_x : Y \to Y$ is a continuous map. Then if $T : X \to X$ is a dynamical system, we define the skew product of T and S as a dynamical system on $X \times Y$:

$$T \ltimes S(x, y) = (T(x), S_x(y))$$

If S is a constant function, then the product is *direct*, and we write this as $T \times S$. We may think of x-coordinate as being very straightforward and decisive with its own future, moving forward in time with the system T. But the y-coordinate as unsure of how to proceed. It therefore asks the x-coordinate what it thinks at each step. But the x-coordinate can be arbitrary, given what state it's in, so the answer y receives will depend on where the x-coordinate sits.

It is clear that $T : X \to X$ is a factor of $T \ltimes S$ using the standard projection from $X \times Y$ to X. In this case we say that $T \ltimes S$ is a semidirect product over T.

EXAMPLE 3.3.1. Let $X, Y = \mathbb{R}/\mathbb{Z}$, $T = R_\alpha$ for some α and $S : \mathbb{R}/\mathbb{Z} \to C(\mathbb{R}/\mathbb{Z}, \mathbb{R}/\mathbb{Z})$ be $S(\beta) = R_\beta$. Then the skew product of T and S is defined by:

$$T \ltimes S(x, y) = (T(x), S_x(y)) = (R_\alpha(x), R_x(y)) = (x + \alpha, y + x)$$

Exercises.

1. Show that if $S : X \to X$ is a factor of $T : Y \to Y$ with factor map $f : Y \to X$, and $\mathcal{O}_T(y)$ is dense in y, then $\mathcal{O}_S(f(y))$ is dense in X.
2. Are the dynamical systems $T_{1/5}$ and $T_{2/5}$ conjugate? What about the dynamical systems $T_{1/3}$ and $T_{2/3}$? Find a necessary and sufficient condition for T_α to be conjugate to T_β, when $\alpha, \beta \in \mathbb{Q}$.
3. Let G be a compact abelian topological group, and $g \in G$. Let $T_g : G \to G$ be the dynamical system defined by $T_g(h) = h + g$.
   (a) Show that $H = \overline{\mathcal{O}(0)}$ is a closed subgroup of G, and that $\overline{\mathcal{O}(x)}$ is a coset of H for every $x \in G$.
   (b) Show that if $G = \mathbb{Z}_p$ and $g = 1$, then every orbit is dense.
   (c) Show that $T_g \times T_g : G \times G \to G \times G$ never has a dense orbit.
4. Prove or find a counterexample: If $T_2 : X_2 \to X_2$ is a factor of $T_1 : X_1 \to X_1$ and $T_3 : X_3 \to X_3$ is a factor of T_2, then T_3 is a factor of T_1.
5. * Prove or find a counterexample: If $T : X \to X$ is a factor of $S : Y \to Y$ and $S : Y \to Y$ is a factor of $T : X \to X$, then S and T are topologically conjugate.

CHAPTER 4

Homotopy Theory and the Fundamental Group

The contents of this section will begin our investigation into algebraic topology. A good (and free!) reference text for this is Allen Hatcher's *Algebraic Topology.*

4.1. Lifting Properties and Degree

Let $T_\alpha : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ be the map $[t] \mapsto [t+\alpha]$. Recall the topological semiconjugacy:

$$\begin{array}{ccc} \mathbb{R} & \xrightarrow{\widetilde{T}_\alpha} & \mathbb{R} \\ p\big\downarrow & & \big\downarrow p \\ \mathbb{R}/\mathbb{Z} & \xrightarrow[T_\alpha]{} & \mathbb{R}/\mathbb{Z} \end{array}$$

Does this work for any map? If $f : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ is a homeomorphism, is there a continuous map $\tilde{f}$ such that:

$$\begin{array}{ccc} \mathbb{R} & \xrightarrow{\tilde{f}} & \mathbb{R} \\ p\big\downarrow & & \big\downarrow p \\ \mathbb{R}/\mathbb{Z} & \xrightarrow[f]{} & \mathbb{R}/\mathbb{Z} \end{array}$$

The answer is yes, but not uniquely. In fact, even in the case of the circle, the map T_α can be lifted to $\widetilde{T}_{\alpha+n}$ for any $n \in \mathbb{Z}$ (Why?). To make this lifting problem more precise, we need to develop some tools of algebraic topology.

Theorem 4.1.1 (Path Lifting Property). *Let $\gamma : [0,1] \to \mathbb{R}/\mathbb{Z}$ be a continuous map, and let $p : \mathbb{R} \to \mathbb{R}/\mathbb{Z}$ be the quotient map. Choose some $x_0 \in p^{-1}(\gamma(0))$. Then there exists a unique continuous $\tilde{\gamma} : [0,1] \to \mathbb{R}$ such that $\tilde{\gamma}(0) = x_0$ and $p \circ \tilde{\gamma} = \gamma$. That is, the following diagram commutes:*

$$\begin{tikzcd} & \mathbb{R} \arrow[d, "p"] \\ {[0,1]} \arrow[ur, "\tilde{\gamma}"] \arrow[r, "\gamma"'] & \mathbb{R}/\mathbb{Z} \end{tikzcd}$$

Before proving the path lifting property, let us see how it solves our lifting problem:

Corollary 4.1.2. *Given a continuous map $f : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$, there exists a continuous map $\tilde{f} : \mathbb{R} \to \mathbb{R}$ such that $p \circ \tilde{f} = f \circ p$*

Proof. Choose $x_0 \in \mathbb{R}$, and begin by defining $\tilde{f}(x_0) = y_0$ for some y_0 such that $p(y_0) \in f(p(x_0))$. Define $\eta_x : [0,1] \to \mathbb{R}$ by $\eta_x(t) = (1-t)x_0 + tx$. Let $\gamma_x : [0,1] \to \mathbb{R}/\mathbb{Z}$ be $\gamma_x = f \circ p \circ \eta_x$. Then by Theorem 4.1.1, there exists a unique $\tilde{\gamma}_x : [0,1] \to \mathbb{R}$ such that $\tilde{\gamma}_x(0) = y_0$ and $p \circ \tilde{\gamma}_x = \gamma_x$. Now define $\tilde{f}(x) = \tilde{\gamma}_x(1)$. By construction:

$$p \circ \tilde{f}(x) = p \circ \tilde{\gamma}_x(1) = \gamma_x(1) = f \circ p \circ \eta_x(1) = f \circ p(x)$$

Continuity is intuitively clear, but can be a bit of a bear in the details. We sketch an outline here.

Fix $x \in \mathbb{R}$ and $\varepsilon > 0$. It suffices to consider the case when $\varepsilon < 1/2$. Because f is continuous, there exists a $\delta > 0$ such that if $d(p(x), p(y)) < \delta$, then $d(f(p(x)), f(p(y))) < \varepsilon$. We assume that $x_0 < x$ (the other case has a symmetric argument). Let $\eta = \eta_{x+\delta} : [0,1] \to \mathbb{R}$ be as before. For $y \in B_\delta(x)$, define $c_y : [0,1] \to [0,1]$ by $c_y(t) = \frac{y-x_0}{x+\delta-x_0}$. Then by direct computation, $\eta_y = \eta \circ c_y$. Similarly, let $\gamma = \gamma_{x+\delta}$. In particular, by uniqueness of lifts, $\tilde{f}(y) = \tilde{\gamma}(c_y(1))$ and $\tilde{f}(x) = \tilde{\gamma}(c_x(1))$. Then:

$$d(\tilde{f}(x), \tilde{f}(y)) = d(\tilde{\gamma}(c_x(1)), \tilde{\gamma}(c_y(1))) = d(\gamma(c_x(1)), \gamma(c_y(1)) = d(f(p(x)), f(p(y)) < \varepsilon$$

The second equality will follow from the constructions of $\tilde{\gamma}$ outlined in the proof of Theorem 4.1.1.

□

Note that we have made no claims about uniqueness in Corollary 4.1.2. A more precise claim will be made later.

Proof of Theorem 4.1.1. The idea of the proof is as follows: $\mathbb{R}/\mathbb{Z}$ is "locally like an open interval." By this, we mean that if you pick a point and a small neighborhood of that point, the neighborhood is homeomorphic to the interval $(0, 1)$. Then, at least locally, we can invert the map p. So instead of lifting γ all at once, where each point can have many preimages, we lift a specified point first, then expand out in a neighborhood of that lift. We continue this way until all of the intervals are lifted.

Let $\gamma : [0,1] \to \mathbb{R}/\mathbb{Z}$ and x_0 be as in the statement of the theorem. We want to define $\tilde{\gamma} : [0,1] \to \mathbb{R}$. Let $\varepsilon > 0$ be such that $p|_{B_\varepsilon(x_0)} : B_\varepsilon(x_0) \to \mathbb{R}/\mathbb{Z}$ is a homeomorhpism onto its image (Why does such a $\varepsilon > 0$ exist?). For notational convenience, let $U = B_\varepsilon(x_0)$ and $p|_U = p_U$. Then there is some $\delta_0 > 0$ such that if $t < \delta_0$, $d(\gamma(0), \gamma(t)) < \varepsilon$. Then we can define $\tilde{\gamma}|_{[0,\delta]} = p_U^{-1} \circ \gamma$. To continue extending the definition of γ to all of $[0, 1]$, we need what is called the covering property.

Proposition 4.1.3 (Covering Property). *There is a cover $\mathcal{U} = \{U_\alpha : \alpha \in A\}$ of $\mathbb{R}/\mathbb{Z}$ such that:*

(i) Each U_α is open in $\mathbb{R}/\mathbb{Z}$
(ii) $\mathbb{R}/\mathbb{Z} = \cup_{\alpha \in A} U_\alpha$
(iii) $p^{-1}(U_\alpha) = \cup_{i \in \mathbb{Z}} V_\alpha^i$
(iv) V_α^i are open sets of $\mathbb{R}$ such that $V_\alpha^i \cap V_\alpha^j = \emptyset$ when $i \neq j$
(v) $p|_{V_\alpha^i} : V_\alpha^i \to U_\alpha$ is a homeomorphism for each i

To see this in the case of $\mathbb{R}/\mathbb{Z}$, we can choose $U_1 = \mathbb{R}/\mathbb{Z} \setminus \{[0]\}$ and $U_2 = \mathbb{R}/\mathbb{Z} \setminus \{[1/2]\}$. The preimages will be $V_1^i = (i, i+1)$ and $V_2^i = ((2i-1)/2, (2i+1)/2)$. The idea for moving forward is now clear, and we will draw a caricature below (Figure 1). Imagine our space is not a circle but a disc, covered by the regions shaded red and blue. Then let γ be the curve depicted in brown. While γ does not lie completely in blue or in red alone we can find subintervals of $[0, 1]$ in which it the restriction does lie in one of them. The idea is to "paste" these restrictions together to form a well-defined map γ.

There is a slight problem, however. What if the curve passes infinitely often between the blue and red regions (ie, the open cover by the blue and red sets)? Luckily for us, infinite oscillation is impossible, because the curve has to accumulate to a *single* point on its endpoints (ie, $[0, 1]$ is compact). If we replaced $[0, 1]$ by $[0, 1)$ we would be out of luck!

Let's look more closely at the details of the argument. Consider the covering $W_1 = \gamma^{-1}(U_1)$ and $W_2 = \gamma^{-1}(U_2)$ of $[0, 1]$. We wish to find a finite, increasing sequence of points $0 = t_0 < t_1 \cdots < t_n = 1$ such that $[t_i, t_{i+1}] \subset W_i$ for some i. By continuity of γ,, the sets W_i are open, we can always find subsequent t_i. A priori, the sequence t_i may not terminate. To see that we can always reach the point 1, we guarantee that there is some δ for which we can insist that $t_i - t_{i-1} \geq \delta$. Such a δ will

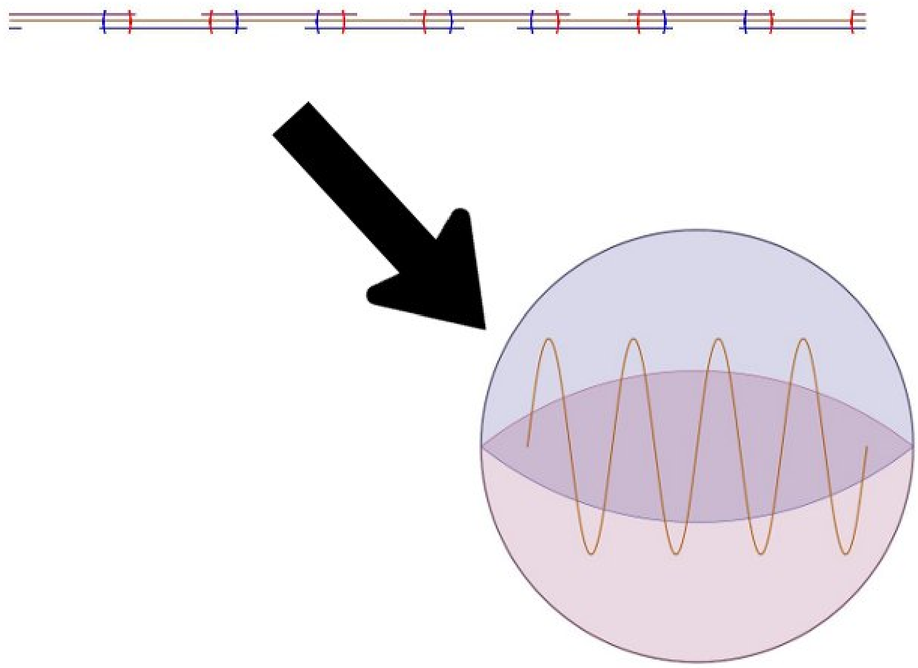

FIGURE 1. A curve in X

be called a *Lesbesgue number* for the covering W_1, W_2 and will be proven to exist below (Lemma 4.1.4).

Then iterating the arguments in the first part of the proof, replacing $[0, \delta]$ with $[t_i, t_{i+1}]$ will give us a lift of $\gamma|_{[t_i,t_{i+1}]}$ for every i. Furthermore, we can insist that the maps match on the endpoints. Thus the map lifts to a continuous map $\tilde{\gamma} : [0, 1] \to \mathbb{R}$.

Let us now prove uniqueness. Suppose that $\tilde{\gamma}_1$ and $\tilde{\gamma}_2$ are two lifts such that $\tilde{\gamma}_1(0) = \tilde{\gamma}_2(0)$ and $p \circ \tilde{\gamma}_1 = p \circ \tilde{\gamma}_2 = \gamma$. Then because p is a homomorphism, $\eta = \tilde{\gamma}_2 - \tilde{\gamma}_1$ is a lift of the constant path at 0. But then $\eta : [0, 1] \to \mathbb{Z}$, becase $\mathbb{Z}$ are the only points which project to 0. One easily observes that the only continuous maps from $[0, 1]$ to $\mathbb{Z}$ are the constant maps (use the intermediate value theorem), so $\tilde{\gamma}_1(t) = \tilde{\gamma}_2(t) + n$ for some fixed n. Since they match at 0, $n = 0$ and we get uniqueness. □

LEMMA 4.1.4 (Lesbesgue Number Lemma). *Let X be a compact metric space, and $\mathcal{U} = \{U_\alpha : \alpha \in A\}$ an open cover of X. Then there exists some $\delta > 0$ such that for every $x \in X$, there exists a $U_\alpha \in \mathcal{U}$ such that $B_\delta(x) \subset U_\alpha$*

PROOF. First, since X is compact, there is an finite subcollection of the U_α such that $X = U_1 \cup U_2 \cup \cdots \cup U_m$. Then let $C_i = X \setminus U_i$, and consider the function $f : X \to \mathbb{R}$ defined by $f(x) = \frac{1}{m} \sum_{i=1}^{m} d(x, C_i)$, where $d(x, C_i) = \inf \{d(x, y) : y \in C_i\}$. Then since for every $x \in X$, $x \in U_i$ for some i, $f(x) > 0$ on all of X. Since X is compact, and its image is in $(0, \infty)$, there must be some δ such that $f(x) \geq \delta$ for every $x \in X$.

Let us now verify the property we want δ to satisfy. Fix $x \in X$ and let j be the index of i for which $d(x, C_i)$ takes its maximal value. Then $\delta < f(x) = \frac{1}{m} \sum_{i=1}^{m} d(x, C_i) \leq \frac{1}{m} \cdot md(x, C_j) = d(x, C_j)$. In other words, $B_\delta(x) \cap C_j = \emptyset$ and $B_\delta(x) \subset U_j$. □

For now, we will be content to work within the circle, but we document a crucial notion we came across in the proof of Theorem 4.1.1 for later abstractions:

Definition 4.1.5. Let X be a metric space. A continuous surjective map $p : Y \to X$ is a *covering map* if p satisfies the properties of Proposition 4.1.3, replacing $\mathbb{R}/\mathbb{Z}$ by X and $\mathbb{R}$ by Y. That is, there is a cover $\mathcal{U} = \{U_\alpha : \alpha \in A\}$ of X such that:

(i) Each U_α is open in X
(ii) $X = \cup_{\alpha \in A} U_\alpha$
(iii) $p^{-1}(U_\alpha) = \cup_{i \in \mathbb{Z}} V_\alpha^i$
(iv) V_α^i are open sets of Y such that $V_\alpha^i \cap V_\alpha^j = \emptyset$ when $i \neq j$
(v) $p|_{V_\alpha^i} : V_\alpha^i \to U_\alpha$ is a homeomorphism for each i

Can we detect different ways to lift a path somehow using the cover $\mathbb{R}$? We can, in fact, and to do so we need to work with some groups:

Definition 4.1.6. Let $\mathcal{H} = \mathcal{H}(\mathbb{R}, \mathbb{R})$ be the space of homeomorphisms from $\mathbb{R}$ to $\mathbb{R}$. Then $\mathcal{H}$ is a group, where the operation is compositions. Let $\Gamma \subset \mathcal{H}(\mathbb{R}, \mathbb{R})$ be the group of *deck transformations* of the cover $p : \mathbb{R} \to \mathbb{R}/\mathbb{Z}$. That is, $f \in \Gamma$ if and only if $p \circ f = p$.

Exercise 4.1.1. Verify that Γ is a subgroup of $\mathcal{H}$

Proposition 4.1.7. *For the covering $p : \mathbb{R} \to \mathbb{R}/\mathbb{Z}$, the group of deck transformations is the group $\Gamma = \{f_n : f_n(x) = x + n\}$*

Proof. We need to prove two inclusions. First note that $p(f_n(x)) = p(x+n) = [x+n] = [x] = p(x)$. Now suppose that $g : \mathbb{R} \to \mathbb{R}$ is a deck transformation. Then $f(x) = g(x) - x$ is a map such that $p \circ f = 0$ (ie, $p \circ f$ is the constant map to 0). Then $f : \mathbb{R} \to \mathbb{Z}$. By the intermediate value theorem, such a map must be constant, and $f(x) \equiv n$. That is, $g(x) = x + n$ for some n. $\square$

Here is an important observation. The group Γ is isomorphic to $\mathbb{Z}$, under the map $\varphi : \mathbb{Z} \to \Gamma$ which sends $n \to f_n$. Let us return to circle maps $f : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ and lift it to a map $\tilde{f} : \mathbb{R} \to \mathbb{R}$. Let us try to simplify the complicated way that we have constructed this lift:

Example 4.1.1.

(1) For $d \in \mathbb{Z}$, define $L_d : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ by $L_d([x]) = [dx]$. Then we can lift L_d to $\widetilde{L_d}(x) = dx$. The only other lifts of this map are of the form $x \mapsto dx + c$ for some $c \in \mathbb{Z}$.
(2) Let $S : \mathbb{R}\ \mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ by $S([x]) = [\sin(2\pi x)]$. Then, a lift of S is $\widetilde{S}(x) = \sin(2\pi x)$. Other lifts take the form $x \mapsto \sin(2\pi x) + n$ for $n \in \mathbb{Z}$.

Here is one nice feature of the second example: S not only lifts to a map $\widetilde{S} : \mathbb{R} \to \mathbb{R}$, but also to a map $\widehat{S} : \mathbb{R}/\mathbb{Z} \to \mathbb{R}$ defined by $\widehat{S}([x]) = \sin(2\pi x)$, but there does not exist such a $\widehat{L_d}$ for $d \neq 0$. Investigating further into the topology of $\mathbb{R}$ and $\mathbb{R}/\mathbb{Z}$ will give us good reason to expect this!

Let us consider a modification of the maps L_d. Let $\tilde{\varphi} : \mathbb{R} \to \mathbb{R}$ be a function of period 1. That is, $\tilde{\varphi}(x+1) = \tilde{\varphi}(x)$ for every $x \in \mathbb{R}$. Then define $\tilde{f} : \mathbb{R} \to \mathbb{R}$ by $\tilde{f}(x) = dx + \tilde{\varphi}(x)$. Then by construction, the function $f([x]) = [\tilde{f}(x)]$ is a well-defined self-map of $\mathbb{R}/\mathbb{Z}$.

Theorem 4.1.8. *If $f : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ is a continuous map, there exists a unique $d \in \mathbb{Z}$ (d is called the degree of f, denoted $\deg f$) such that any lift of f takes the form $\widetilde{L}_d + \tilde{\varphi}$, where $\tilde{\varphi}$ is a 1-periodic map. The maps $\tilde{\varphi}$ are lifts of a unique map $\varphi : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$.*

Let us make a few observations. First, since $\varphi(\mathbb{R}) = \varphi([0,1])$, φ is bounded. As a corollary, we get that there exists a C such that $dx - C \leq \tilde{f}(x) \leq dx + C$. To prove the theorem, we need a Lemma:

Lemma 4.1.9. *Let $f : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ be continuous, then:*

(a) If $\tilde{f}_1 : \mathbb{R} \to \mathbb{R}$ and $\tilde{f}_2 : \mathbb{R} \to \mathbb{R}$ are both lifts of f, then $\tilde{f}_1(x) = \tilde{f}_2(x) + n$ for some constant n
(b) If $\tilde{f} : \mathbb{R} \to \mathbb{R}$ is a lift of f, then $h : \mathbb{R} \to \mathbb{R}$ defined by $h(x) = \tilde{f}(x+1)$ is also a lift

PROOF. These claims follow easily from the definition of a lift of f. Namely, $\tilde{f}$ is a lift if and only if $\tilde{f}$ is continuous and $p \circ \tilde{f} = f \circ p$. That is, the following diagram commutes:

$$\begin{array}{ccc} \mathbb{R} & \xrightarrow{\tilde{f}} & \mathbb{R} \\ {\scriptstyle p}\big\downarrow & & \big\downarrow{\scriptstyle p} \\ \mathbb{R}/\mathbb{Z} & \xrightarrow[f]{} & \mathbb{R}/\mathbb{Z} \end{array}$$

We leave the first claim as an exercise. To see the second claim, note that $p \circ h(x) = p \circ \tilde{f}(x+1) = f([x+1]) = f([x])$ and $f \circ p(x) = f([x])$. ☐

Thus for any lift, there exists $d \in \mathbb{Z}$ such that $\tilde{f}(x+1) - \tilde{f}(x) \equiv d$.

PROOF OF THEOREM 4.1.8. Let $\tilde{f}_0$ be a lift of f. Then $d \equiv \tilde{f}_0(x+1) - \tilde{f}_0(x) \in \mathbb{Z}$. Let $\tilde{\varphi}_0 : \mathbb{R} \to \mathbb{R}$ by $\tilde{\varphi}_0(x) = f_0(x) - dx$. Then by construction, $\tilde{\varphi}_0(x+1) = f_0(x+1) - d(x+1) = f_0(x) + d - dx - d = f_0(x) - dx = \tilde{\varphi}_0(x)$.

We need only to show uniqueness. Let $\tilde{f} : \mathbb{R} \to \mathbb{R}$ be any other lift of f. Then there exists $n \in \mathbb{Z}$ such that $\tilde{f}(x) = f_0(x) + n$, so that $\tilde{f}(x) = dx + (\tilde{\varphi}_0(x) + n)$. Thus $\tilde{\varphi}$ is defined uniquely up to an integer, and hence corresponds to a unique map $\varphi : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$. ☐

Note that because of the power of existence and uniqueness of lifts, the proof took a simple form. One immediate observation is that $\deg(L_d) = d$.

COROLLARY 4.1.10. *Given $g : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ a continuous map, there exists a map $\hat{g} : \mathbb{R}/\mathbb{Z} \to \mathbb{R}$ such that $p \circ \hat{g} = g$ if and only if* $\deg(g) = 0$

PROOF. Suppose that such a function $\hat{g} : \mathbb{R}/\mathbb{Z} \to \mathbb{R}$ exists. Then the function $\tilde{g} = \hat{g} \circ p$ is a lift of g. It is 1-periodic, and hence takes the form $\tilde{g}(x) = 0 \cdot x + \tilde{g}(x)$. So $\deg(g) = 0$.

Now suppose that $\deg(g) = 0$. Then there exists a lift $\tilde{g}$ which is 1-periodic. This allows us to define $\hat{g}(x) = \tilde{g}([x])$. This is well-defined because $\tilde{g}(x + \mathbb{Z}) = \tilde{g}(x)$. By construction, $\hat{g}$ satisfies the desired property. ☐

4.2. Brouwer's Fixed Point Theorem and Homotopy

Let us think about maps of the interval:

THEOREM 4.2.1. *Let $f : [0,1] \to [0,1]$ be a continuous map. Then f has a fixed point.*

PROOF. Note that $f(0) \geq 0$ and $f(1) \leq 1$, so that $h(x) = f(x) - x$ has $h(0) \geq 0$ and $h(1) \leq 0$. By the Intermediate Value Theorem, $h(x) = 0$ for some x. That is, $f(x) = x$ ☐

Let us try to find another way to prove this. The ideas of this proof will help us with generalizations:

SECOND PROOF. Assume there does not exist a fixed point. Let $h : [0,1] \to \{0,1\}$ be defined by $h(x) = 0$ if $x > f(x)$ (ie, the ray starting at x and passing through $f(x)$ hits 0) and $h(x) = 1$ if $x < f(x)$ (this has a similar characterization using rays). Then h is continuous, since x can be separated from $f(x)$ by an open set. Thus, either $h([0,1]) = \{0\}$ and $h([0,1]) = \{1\}$. But this is impossible, since $f(0) > 0$ and $f(1) < 1$ ☐

The interval in $\mathbb{R}$ acts like a ball in $\mathbb{R}^2$, so we should not be suprised by:

THEOREM 4.2.2 (Brouwer's Fixed Point Theorem). *Let $\overline{\mathbb{D}} = \{z \in \mathbb{C} : |z| \leq 1\}$. Let $f : \overline{\mathbb{D}} \to \overline{\mathbb{D}}$ be continuous. Then f has a fixed point.*

Let us try to naively adapt the Second Proof of the 1-dimensional version to this case:
Proof. Let $h : \overline{\mathbb{D}} \to S^1$ be defined by the following rule: $h(x)$ is the point on S^1 which the ray starting at $f(x)$ and passing through x hits. See Figure 2. Then:

1. h is continuous
2. $h|_{S^1} = \mathrm{id}$

The continuity of h is intuitively clear, but a rigorous proof is posted as an auxiliary note on the website. So we will have the proof once we have:

PROPOSITION 4.2.3. *There does not exist a function $h : \overline{\mathbb{D}} \to S^1$ such that $h|_{S^1} = \mathrm{id}$*

We need some more tools to prove this, but intuitively, we can't pull the disc to its boundary without puncturing it somewhere! Imagine we had a vinyl record. We want to map the entire record to the boundary, so we start taking the inner circles and expanding them out. This all works well and fine until we get to the center! If we wanted to define where the very middle of the record went, we'd be out of luck. Interestingly enough, there do exist maps $h : \overline{\mathbb{D}} \setminus B_{1/4}(0) \to S^1$ which are the identity on the boundary, so in the case of a vinyl record (the corresponding topological space is called the *annulus*), when we don't need to define the map at the center, we're actually OK! See Figure 3.

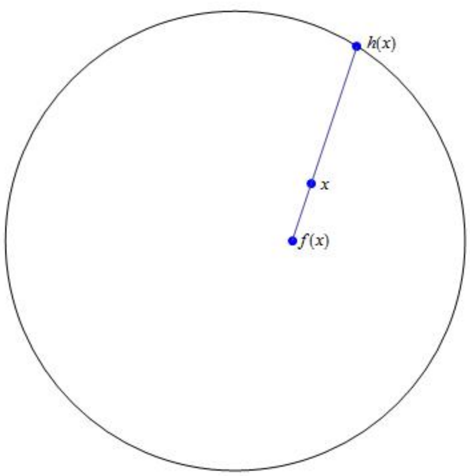

FIGURE 2. How the map h is constructed

FIGURE 3. Vinyl Records and the Topology of the Disc

EXERCISE 4.2.1. Show that under the metric $\hat{d}$, $(x_n, y_n) \to (x, y)$ if and only if $x_n \to x$ and $y_n \to y$

To prove Proposition 4.2.3, we will use *polar coordinates*. Namely, we will exploit the map $P : [0, 1] \times \mathbb{R}/\mathbb{Z} \to \overline{\mathbb{D}}$ defined by $P(r, \theta) = r \exp(2\pi i \theta) = r \cos\theta + r \sin\theta$. The map P can be thought of as taking a cylinder, and collapsing one of the ends to the origin (see Figure 4).

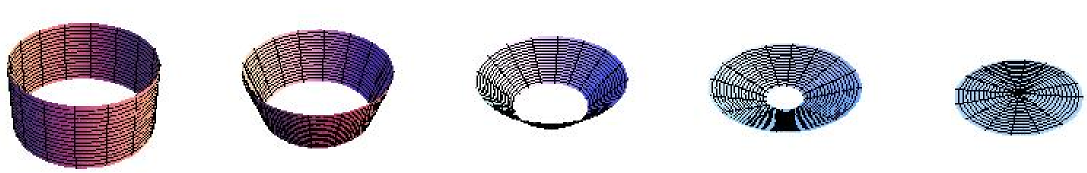

FIGURE 4. Visualizing $P : [0, 1] \times \mathbb{R}/\mathbb{Z} \to \overline{\mathbb{D}}$

Given the map h, we can construct $R : [0,1] \times \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ defined by $R(r,\theta) = E^{-1} \circ h \circ P$. Since $h|_{S^1} = \mathrm{id}$, we know that $R(1,\theta) = \theta$, computing that:

$$R(1,\theta) = E^{-1}(h(P(1,\theta))) = E^{-1}(h(\exp(2\pi i\theta))) = E^{-1}(\exp(2\pi i\theta)) = \theta$$

Observe also that $R(0,\theta) = E^{-1}(h(0))$, so R_0 is a constant map. R is a special type of map, known as a *homotopy*.

Definition 4.2.4. Let X and Y be topological spaces, and $f, g : X \to Y$ be continuous maps. A *homotopy* (between f and g) is a continuous map $H : [0,1] \times X \to Y$ such that:

$$\begin{aligned} f(x) &= H(0,x) \\ g(x) &= H(1,x) \end{aligned}$$

We will often denote $H(t,x)$ by $H_t(x)$. We think of H as a continuously varying family of maps connecting f and g

Exercise 4.2.2. Consider the relation $f \sim g$ if and only if there exists a homotopy between f and g. Show that $\sim$ is an equivalence relation. If $f \sim g$, then f is said to be *homotopic* to g.

R is then a homotopy between a constant map and the identity on $\mathbb{R}/\mathbb{Z}$. So we can reformulate Proposition 4.2.3 as the following:

The map $\mathrm{id} : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ is not homotopic to a constant map

We can prove this from the following theorem:

Theorem 4.2.5 (Homotopy Lifting Property). *Let $H : [0,1] \times Y \to \mathbb{R}/\mathbb{Z}$ be a homotopy, and suppose that there exists a map $\tilde{h} : Y \to \mathbb{R}$ such that $p \circ \tilde{h} = H_0$. Then if p is the canonical projection, there exists a unique lift $\widetilde{H} : [0,1] \times Y \to \mathbb{R}$ such that $p \circ \widetilde{H} = H$ and $\widetilde{H}_0 = \tilde{h}$.*

Let us finish the proof of the Brouwer Fixed Point Theorem. Assume, for a contradiction, that there existed a homotopy $H : [0,1] \times \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ such that H_0 is a constant map, and $H_1 = \mathrm{id}$. Since H_0 is a constant map, it has degree 0, and hence a lift $\widetilde{H}_0 : \mathbb{R}/\mathbb{Z} \to \mathbb{R}$. Thus there must exist a lift of the homotopy $\widetilde{H}$. But then $\widetilde{H}_1$ is a lift of $H_1 = \mathrm{id}$. But by Corollary 4.1.10, such a lift exists if and only if $\deg(H_1) = 0$. Since $\deg(H_1) = \deg(\mathrm{id}) = 1$, we have reached a contradiction.

Before we prove Theorem 4.2.5, we need to observe the following properties of the product space $X \times Y$ which will be left as exercises:

(i) If $U \subset X$ and $V \subset Y$, then $U \times Y$ is open in $X \times Y$
(ii) If $W \subset X \times Y$ is open, and $z \in W$, there exists open sets $U \subset X$ and $V \subset Y$ such that $z \in U \times V \subset W$
(iii) $\{x\} \times Y$ is a closed subset of $X \times Y$, and the map $i_x : Y \to X \times Y$ given by $i_x(y) = (x,y)$ is a homeomorphism onto its image

Lemma 4.2.6. *Let Y be a metric spce, and $\{W_j \subset [0,1] \times Y : j \in J\}$ be an open cover of $[0,1] \times \{y\} \subset [0,1] \times Y$. Then there exists an open set $N \subset Y$ such that $y \in N$ and a partition $0 = t_0 < t_1 < \cdots < t_n = 1$ of $[0,1]$ such that for every $i = 0, \ldots, n-1$, there exists some $j \in J$ such that $[t_i, t_{i+1}] \times N \subset W_j$*

Let us make some final observations before entering the proof of Theorem 4.2.5. Let $y \in Y$. Then the map $t \mapsto \widetilde{H}(t,y)$ will be a path in $\mathbb{R}$. Note that the starting point $\widetilde{H}(0,y) = \tilde{h}(y)$ is already given. Thus we are left with only one option: these paths will have to be lifts of the corresponding paths in $\mathbb{R}/\mathbb{Z}$. The only thing we have to verify is that if we lift path-by-path, the lift will be a continuous map! Unfortunately, proving continuity will amount essentially to reproving the path lifting lemma, as we shall see.

Lemma 4.2.7. *Let $y \in Y$ and $\tilde{h} : Y \to \mathbb{R}$ be a lift of $h = H_0$. Then there exists an open N such that $y \in N \subset Y$, and a unique lift $\widetilde{H}|_N : [0,1] \times N \to \mathbb{R}$ such that $\widetilde{H}|_N(0,y) = \tilde{h}|_N(y)$.*

Proof. Recall Proposition 4.1.3, and let U_α be an open covr of $\mathbb{R}/\mathbb{Z}$ with the perscribed properties. Now let $W_\alpha = H^{-1}(U_\alpha)$ be the preimages of U_α, so that W_α is an open cover of $[0,1]\times Y$. By Lemma 4.2.6. there is a neighborhood N of $y \in Y$ and numbers $0 = t_0 < t_1 < \cdots < t_n = 1$ such that $[t_i, t_{i+1}] \times N \subset W_\alpha$ for some α. We proceed now by induction on n. In the base case $n = 1$. Note that by construction, $\tilde{h}([t_0, t_1] \times N)$ contained in a unique V_{α_0} such that $p|_{V_{\alpha_0}} : V_{\alpha_0} \to U_{\alpha_0}$ is a homeomorphism, so that $q = (p|_{V_{\alpha_0}})^{-1}$ exists and is continuous. Then let $\widetilde{H}(t,y) = q \circ H(t,y)$ for $y \in N$ and $t \in [t_0, t_1]$.

Note that in the above argument, we can replace $[t_0, t_1]$ by any $[t_i, t_{i+1}]$, but we must specify the map $\tilde{h}$. By induction, $\tilde{h}_i(y) = \widetilde{H}(t_i, y)$ is a well-defined continuous map, So we can define a lift $\widetilde{H}$ on the set $[t_i, t_{i+1}]$ such that $\widetilde{H}(t_i, y) = \tilde{h}_i(y)$.

$\square$

Proof of Theorem 4.2.5. Cover Y be the neighborhoods as appearing in Lemma 4.2.7. That is, $Y = \cup_{y\in Y} N_y$, where N_y is the neighborhood as in the Lemma, and let G_y be the local homotopy that we construct which matches with $\tilde{h}$ on N_y. Define $\widetilde{H}(t,y) = G_y(t,y)$. We claim that $\widetilde{H}$ is continuous. To see this, we will show that if $N_y \cap N_z \neq \emptyset$, then $G_y = G_z$ on this intersection. This follows because if $w \in N_y \cap N_z$, both $t \mapsto G_y(t,w)$ and $t \mapsto G_z(t,w)$ are both lifts of $t \mapsto H(t,w)$ such that $G_y(0,w) = G_z(0,w) = \tilde{h}(w)$. By the uniqueness of path lifts (Theorem 4.1.1), we get that $G_y(t,w) = G_z(t,w)$. We get continuity from the following final lemma below.

$\square$

Lemma 4.2.8 (Pasting Lemma). *Let $f : Z \to X$ be a function (not a priori continuous) on metric spaces, and $Z = \cup_{\alpha\in A} U_\alpha$ with U_α open. Then f is continuous if and only if $f|_{U_\alpha} : U_\alpha \to X$ is continous for every α.*

Proof. Let V be open in X. Then $f^{-1}(V) = \cup_{\alpha\in A} U_\alpha \cap f^{-1}(V)$. But since $f|_{U_\alpha}$ is continuous, $f|_{U_\alpha}^{-1}(V) = U_\alpha \cap f^{-1}(V)$ is open, and hence the union of such sets over all α is open. So $f^{-1}(V)$ is open and f is continuous. $\square$

Many of the results we found actually hold in much greater generality, the generality of covering spaces (Definition 4.1.5).

Example 4.2.1.

(1) $\mathrm{id} : X \to X$ is a covering map. In this case, the cover of X is just $\mathcal{U} = \{X\}$ and $\mathrm{id} : X \to X$ is a homeomorphism.
(2) Let X be a metric space, and $Y = X \times \{0,1\}$. Let $p : Y \to X$ be $p(x,t) = x$. Again, we can use the cover $\mathcal{U} = \{X\}$, and this time, $p^{-1}(X) = X \times \{0\} \sqcup X \times \{1\}$, and p restricted to $X \times \{n\}$ is a homeomorphism for $n = 0, 1$.
(3) We can extend the previous example by setting $Y = X \times D$, where D is any discrete space. Then we can use the same cover, and p restricted to $X \times \{d\}$ is a homeomorphism for each $d \in D$
(4) We have seen that $p : \mathbb{R} \to \mathbb{R}/\mathbb{Z}$ is a covering map, let us construct another covering map of $\mathbb{R}/\mathbb{Z}$. We claim that $L_d : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ is a covering map for $d \in \mathbb{Z}$. Indeed, cover $\mathbb{R}/\mathbb{Z}$ be the sets $\mathcal{U} = \{U_1, U_2\}$, where $U_1 = \mathbb{R}/\mathbb{Z} \setminus \{[0]\}$ and $U_2 = \mathbb{R}/\mathbb{Z} \setminus \{1/2\}$. Then $L_d^{-1}(U_1) = \bigsqcup_{i=0}^{d-1} (\frac{i}{d}, \frac{i+1}{d})$ and $L_d^{-1}(U_2) = \bigsqcup_{i=0}^{d-1} (\frac{i}{d} + \frac{1}{2d}, \frac{i+1}{d} + \frac{1}{2d})$

CHAPTER 5

Expanding Maps and Symbolic Dynamics

Let us focus on the case of $L_d : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$, and think about this dynamically.

THEOREM 5.0.9. *Any map $f : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ of degree d has at least $d-1$ fixed points*

PROOF. Note that $x \in \mathbb{R}/\mathbb{Z}$ is fixed by f if and only if $(f - L_1)(x) = [0]$. Let $\tilde{g}$ be a lift of $f - L_1$, so that $\tilde{g}(1) - \tilde{g}(0) = d-1$. Then $(f - L_1)([t]) = [0]$ if and only if $(f - L_1)(t) \in \mathbb{Z}$. We proceed with the case when $d - 1 \geq 0$ (the case of $d - 1 < 0$ follows similarly). Since $d - 1 \geq 0$, we know that $\tilde{g}([0,1)) \supset [\tilde{g}(0), \tilde{g}(1))$. But in any half-open interval of length $d-1$, there exist exactly $d-1$ distinct integer points. Then by the intermediate value theorem, there are real numbers $t_1, \ldots, t_{d-1} \in [0,1)$ which are mapped to distinct integers. This implies that they are all distinct, and since no pair of real numbers $t, s \in [0,1)$ have $[t] = [s]$ and $t \neq s$, we know that $[t_1], \ldots, [t_{d-1}]$ are all distinct fixed points of f. $\square$

COROLLARY 5.0.10. *If $f : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ has degree d, $\left|\mathrm{Fix}(f^k)\right| \geq d^k - 1$*

PROOF. This amount to observing that $\deg(f^k) = \deg(f)^k$, and applying Theorem 5.0.9. $\square$

Here are some natural questions:

1. When is f topologically transitive?
2. Is there a point x such that $\omega(x)$ is a proper subset of $\mathbb{R}/\mathbb{Z}$ which is not periodic?

5.1. Shift Spaces and Symbolic Covers

To answer these questions, it will be useful to think about interval dynamics. After all, we went through great pains to think about the circle as the interval $[0,1]$, identifying 0 with 1. Consider the map $T_3 : [0, 1/3] \cup [2/3, 1] \to [0,1]$ defined by $T_3(x) = 3x \pmod 1$. Note that T_3 is *not* a dynamical system! The domain and codomain of the map is not the same space!

How can we fix this problem? We need to "remove the bad points." For instance, note that $T_3^2(x)$ is well-defined if and only if $T_3(x) \in [0, 1/3] \cup [2/3, 1]$, or if $x \notin T_3^{-1}(1/3, 2/3) = (1/9, 2/9) \cup (7/9, 8/9)$. In a similar manner, we can define $T_3^3(x)$ if and only if $T_3^2(x) \in [0, 1/3] \cup [2/3, 1]$ and T_3^2 is well-defined. That is $x \notin T_3^{-2}(1/3, 2/3) \cup T_3^{-1}(1/3, 2/3)$. In general, we can define the k^{th} iterate of T if and only if $x \notin \bigcup_{i=0}^{k} T_3^{-k}(1/3, 2/3)$. Thus we are motivated to define:

$$\Lambda = [0,1] \setminus \bigcup_{i=0}^{\infty} T_3^{-k}(1/3, 2/3) = \bigcap_{i=0}^{\infty} \big([0,1] \setminus T_3^{-k}(1/3, 2/3)\big)$$

There are a few things we note. First, $0 \in \Lambda$, so $\Lambda \neq \emptyset$. Furthermore, $x \in \Lambda$ if and only if $T^k(x) \in [0, 1/3] \cup [2/3, 1]$ for every $k \geq 0$. It is then obvious, that $T_3(x) \in \Lambda$, so $T_3(\Lambda) = \Lambda$. That is, $T_3 : \Lambda \to \Lambda$ *is* a dynamical system.

To further investigate the structure of Λ, we need to learn about ternary expansion. We are familiar with decimal expansion of real numbers. That is, given $x \in [0,1]$, we can wrie $x = 0.x_1x_2x_3\ldots$, where $x_i \in \{0, 1, 2, \ldots, 9\}$. Here, we know that $x = \sum_{i=1}^{\infty} x_i \cdot 10^{-i}$. If $x_1 = 3$, this means that $x \in [3/10, 4/10]$. If $x_2 = 2$, then $x \in [32/100, 33/100]$. This is equivalent to saying that $10 \cdot x - \lfloor 10 \cdot x \rfloor \in [2/10, 3/10]$.

We can do the same procedure with any base.

THEOREM 5.1.1. *Let $d \in \mathbb{N}$ be a fixed with $d \geq 2$. Then for any $x \in [0,1)$, $x = \sum_{k=1}^{\infty} a_k d^{-k}$, where $a_k \in \{0, \ldots, d-1\}$. This expression is unique unless there exists an $N \in \mathbb{N}$ such that $a_k = 0$ for every $k \geq N$ or $a_k = d-1$ for every $k \geq N$*

We will not prove this, but let us comment on uniqueness. Note that if $a_k = d-1$ for $k \geq N$ and $a_{N-1} \neq d-1$:

$$\sum_{k=1}^{\infty} a_k d^{-k} = \sum_{k=1}^{N-1} a_k d^{-k} + \sum_{k=N}^{\infty} a_k d^{-k} = \sum_{k=1}^{N-1} a_k d^{-k} + d^{-N} \sum_{k=0}^{\infty} \frac{d-1}{d^k}$$
$$= \sum_{k=1}^{N-1} a_k d^{-k} + d^{-N}(d-1) \cdot \frac{1}{1-1/d} = \sum_{k=1}^{N-2} a_k d^{-k} + (a_k+1) d^{-(N-1)}$$

So we can express $x = 0.x_1x_2x_3\ldots$, with $x_i \in \{0,1,2\}$. Indeed, this is shorthand for saying $x = \sum_{i=1}^{\infty} x_i 3^{-i}$.

PROPOSITION 5.1.2. *$x \in \Lambda$ if and only if the ternary expansion can take the form $x = 0.x_1x_2x_3\ldots$ with $x_i \in \{0,2\}$. Such ternary expansions are unique.*

Let us push the proof of this proposition until later.

COROLLARY 5.1.3. *Λ is uncountable*

PROOF. We make use of Cantor's classical diagonalization argument. Suppose Λ were enumerable, and write out the ternary expansions $x^{(i)} = 0.x_1^{(i)}x_2^{(i)}x_3^{(i)}\ldots$. Given $a \in \{0,2\}$, let $\bar{a} = 2$ if $a = 0$ and $\bar{a} = 0$ if $a = 2$. That is, $\bar{0} = 2$ and $\bar{2} = 0$. Then set $y = 0.\bar{x}_1^{(1)}\bar{x}_2^{(2)}\bar{x}_3^{(3)}\ldots$. Then it is clear that $y \neq x^{(i)}$ for any i but $y \in \Lambda$. Hence, Λ is not enumerable. ☐

The tools we have developed branch into a whole new area of dynamics, called *symbolic dynamics*:

DEFINITION 5.1.4. Let $\mathcal{A} = \{a_1, \ldots, a_n\}$ be a finite alphabet (eg, $\mathcal{A} = \{0,1,2\}$). $\Sigma_{\mathcal{A}}^{+} = \mathcal{A}^{\mathbb{N}}$ is the *shift space* on the alphabet $\mathcal{A}$. Points in the shift space represent an infinite sequence in $\mathcal{A}$ indexed by $\mathbb{N}$. We will denote Σ_d^{+} be the space space when $\mathcal{A} = \{0,1,\ldots,d-1\}$

Let us make a few notational conventions. Points in $\Sigma_{\mathcal{A}}^{+}$ will be denoted in bold, ie $\mathbf{x} \in \Sigma_{\mathcal{A}}^{+}$. The n^{th} entry of $\mathbf{x}$ will be denoted x_n. A sequence of points in $\Sigma_{\mathcal{A}}^{+}$ will be indexed in the superscript, ie, $\mathbf{x}^{(n)}$. Then $x_k^{(n)}$ will denote the k^{th} entry of the $\mathbf{n}^{\text{th}}$ element of the sequence.

Theorem 5.1.1 allows us to construct an onto map $h : \Sigma_d^{+} \to [0,1]$ by $h(\mathbf{x}) = \sum_{k=1}^{\infty} x_k d^{-k}$. We wish to introduce a topology on Σ_d^{+} which makes h continuous. The following class of metrics do the trick:

$$d_\beta(\mathbf{x},\mathbf{y}) = \begin{cases} \beta^{N(\mathbf{x},\mathbf{y})} & \mathbf{x} \neq \mathbf{y} \\ 0 & \mathbf{x} = \mathbf{y} \end{cases}$$

Here, $\beta \in (0,1)$ and $N(\mathbf{x},\mathbf{y}) = \min\{n : x_n \neq y_n\}$. Let us make a few observations. First, $d(\mathbf{x},\mathbf{y}) = \beta^n$ if and only if they agree on their first n terms. So the longer two sequences coincide, the closer they are in the metric d_β (since $\beta \in (0,1)$). Indeed, we have the following proposition:

PROPOSITION 5.1.5. *A sequence $\mathbf{x}^{(n)} \to \mathbf{x}$ if and only if for every N, there exists n_0 such that $x_k^{(n)} = x_k$ for every $k = 1,2,\ldots,N$ and $n \geq n_0$. That is, $N(\mathbf{x}^{(n)},\mathbf{x}) \geq N$*

PROOF. First, suppose that we have the described convergence condition. Let $\varepsilon > 0$. Choose $N > 0$ such that $\beta^{-N} < \varepsilon$. Then if n_0 is the associated natural number and $k \geq n_0$, $N(\mathbf{x}^{(k)},\mathbf{x}) \geq N$. That is, $d_\beta(\mathbf{x}^{(k)},\mathbf{x}) < \varepsilon$.

Now suppose that $\mathbf{x}^{(k)} \to \mathbf{x}$. Let $N \in \mathbb{N}$ and $\varepsilon = \beta^{-N}$. Let n_0 be such that if $k \geq n_0$, then $d(\mathbf{x}^{(k)},\mathbf{x}) < \varepsilon = \beta^{-N}$. This implies that $N(\mathbf{x}^{(k)},\mathbf{x}) \geq N$. That is, $x_i^{(k)} = x_i$ for $i = 1,\ldots,N$. ☐

Let us define some convenient open sets. In fact, they are just open balls around a point $\mathbf{x} \in \Sigma_d^{+}$. Let $C(a_1,\ldots,a_n) = \left\{\mathbf{x} \in \Sigma_d^{+} : x_i = a_i \ \forall i = 1,\ldots,n\right\}$. Such sets are called *cylinder sets*.

EXERCISE 5.1.1. Show that $C(a_1, \dots, a_n) = B_\varepsilon(\mathbf{x})$ for some $\mathbf{x} \in \Sigma_d^+$ and $\varepsilon > 0$. Use this to show that d_β and d_α induce the same topology on Σ_d^+

Let us make an observation. We claim that there exists no nonconstant continuous map $\gamma : [0,1] \to \Sigma_d^+$. Indeed, if such a map existed, then the following map would be continuous: $t \mapsto d_\beta(\gamma(0), \gamma(t))$. But the range of the function d_β is $\left\{\beta^N : N \geq 1\right\} \cup \{0\}$. There exists no such nonconstant function by the Intermediate Value Theorem, hence $d_\beta(\gamma(0), \gamma(t)) \equiv 0$ and γ is a constant map.[1]

EXERCISE 5.1.2. Show that every cylinder set is both open and closed

PROPOSITION 5.1.6. *Σ_d^+ is compact*

PROOF. Let $\mathbf{x}^{(n)}$ be a sequence in Σ_d^+. Let us construct nested subsequences. Note that $x_1^{(n)} \in \{0, \dots, d-1\}$ for every n. Thus, at least one of the $i \in \{0, \dots, d-1\}$ occurs infinitely many times. That is there is a subsequence $n_{k,1}$, indexed by k such that $x_1^{(n_{k,1})} = a_1$ for some fixed a_1 and every $k \in \mathbb{N}$. We proceed by induction, noting that we can find a subsequence $n_{k,2}$ of $n_{k,1}$ such that $x_2^{(n_{k,2})} = a_2$ for some fixed k_2 and $k \in \mathbb{N}$. Then let $m_k = n_{k,k}$. We claim that $\mathbf{x}^{(m_k)} \to \mathbf{x}$, where $\mathbf{x} = a_1 a_2 a_3 \dots$. Indeed, if $k \geq N$, then $\mathbf{x}^{(m_k)}$ agrees with $\mathbf{x}$ on the first k elements of the sequence by construction. So by Proposition 5.1.5, $\mathbf{x}^{(m_k)} \to \mathbf{x}$. □

We are finally ready to prove:

PROPOSITION 5.1.7. *The map $h : \Sigma_d^+ \to [0,1]$ is continuous with respect to the metric d_β*

PROOF. Let $\varepsilon > 0$. Note that if $\mathbf{x}, \mathbf{y} \in \Sigma_d^+$:

$$|h(\mathbf{x}) - h(\mathbf{y})| = \left| \sum_{k=1}^{\infty} \frac{x_k - y_k}{d^k} \right| \leq \sum_{k=1}^{\infty} d^{-k} |x_k - y_k|$$

Note that since $\sum_{k=1}^{\infty} \frac{d-1}{d^k} < \infty$, $\sum_{k=N}^{\infty} \frac{d-1}{d^k} \to 0$ as $N \to \infty$. Choose N large enough so that this sum is less than ε, and let $\delta = \beta^{-N}$. Then if $d_\beta(\mathbf{x}, \mathbf{y}) < \delta$, $N(\mathbf{x}, \mathbf{y}) \geq N$, so:

$$|h(\mathbf{x}) - h(\mathbf{y})| \leq \sum_{k=1}^{\infty} \frac{|x_k - y_k|}{d^k} \leq \sum_{k=N}^{\infty} \frac{d-1}{d^k} < \varepsilon$$

Thus, h is continuous. □

We are now ready to prove Proposition 5.1.2, in fact, something stronger: that Λ is homeomorphic to a shift space $\Sigma_{\mathcal{A}}^+$.

PROOF OF PROPOSITION 5.1.2. Consider the space Σ_3^+ and the subspace

$$\Sigma_{\{0,2\}}^+ = \left\{ \mathbf{x} \in \Sigma_3^+ : x_k \in \{0,2\} \, \forall k \in \mathbb{N} \right\}.$$

Then $\Sigma_{\{0,2\}}^+$ is compact, since it is a closed subset of a compact metric space, so if we can show that $h|_{\Sigma_{\{0,2\}}^+}$ is a bijective continuous map, it is in fact a homeomorphism.

First, let us show it is surjective. Suppose that $x \in \Lambda$. Then $T^k(x) \in [0,1/3] \cup [2/3,1]$ by definition of Λ. Define $a_k = 0$ if $T^k(x) \in [0,1/3]$ and $a_k = 2$ if $T^k(x) \in [2/3,1]$. Then $\mathbf{a}$ is an element of $\Sigma_{\{0,2\}}^+$. By construction, $h(\mathbf{a}) = x$.

Now let us show injectivity. Suppose that $h(\mathbf{x}) = h(\mathbf{y})$. By Theorem 5.1.1, if $\mathbf{x} \neq \mathbf{y}$, then for some N, $x_k = y_k$ when $k \leq N$, $x_N \neq 0$, $y_N \neq 2$, $x_k = 0$ for $k \geq N+1$ and $y_k = 2$ for $k \geq N+1$. In this case, however, $x_N = y_N + 1$. Since $x_N \neq 0$, we know that $x_N = 1$ or 2. But if x_N is 2, y_N is 1. Hence at least one of the $\mathbf{x}$ or $\mathbf{y}$ has a 1 appearing. That is, h restricted to $\Sigma_{\{0,2\}}^+$ is injective. □

[1]One may recognize that this implies Σ_d^+ is not *path connected*. Indeed, the path connected (and, in fact connected) components are just singleton sets.

We are now ready to establish the symbolic coding for the full map $L_3 : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$. With the language well-developed, things will follow swiftly. Indeed we define the *Shift Map* on Σ_3^+ by $\sigma : \Sigma_3^+ \to \Sigma_3^+$ by $(\sigma(\mathbf{x}))_n = x_{n+1}$. That is, we have the following picture:

$$(x_1, x_2, x_3, \dots) \xrightarrow{\quad\sigma\quad} (x_2, x_3, x_4, \dots)$$

PROPOSITION 5.1.8. *The map $h : \Sigma_3^+ \to \mathbb{R}/\mathbb{Z}$ is a semiconjugacy between $L_3 : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ and $\sigma : \Sigma_3^+ \to \Sigma_3^+$. That is, the following diagram commutes:*

$$\begin{array}{ccc} \Sigma_3^+ & \xrightarrow{\sigma} & \Sigma_3^+ \\ h\downarrow & & \downarrow h \\ \mathbb{R}/\mathbb{Z} & \xrightarrow{L_3} & \mathbb{R}/\mathbb{Z} \end{array}$$

Furthermore, $y = h(\mathbf{x})$ is the unique point in $\mathbb{R}/\mathbb{Z}$ such that ${L_3}^n(y) \in [x_i/3, (x_i + 1)/3]$

Indeed, the proof of the proposition amounts to straighforward observations of our discussion in this section.

5.2. Markov Partitions

To extend our results from the previous section to a more general setting, we require a new notion:

DEFINITION 5.2.1. Let $f : X \to [0, 1]$ be a continuous map such that X is a union of (possibly overlapping) closed subintervals of $[0, 1]$. A *Markov partition* of X with respect to f is a finite collection of closed intervals $\mathcal{P} = \{I_0, I_1, \dots, I_{d-1}\}$ such that:

(i) $X = \cup_{i=0}^{d-1} I_i$
(ii) The intersection of any two intervals is at most the intersection of their boundaries
(iii) $f|_{I_i}$ is continuous and injective
(iv) (Markov Property) For every i, j, either $f(\mathrm{Int}(I_i)) \cap I_j = \emptyset$ or $f(I_i) \supset I_j$

If for every pair i, j, $f(I_i) \supset I_j$, $\mathcal{P}$ is called a *full Markov partition.* If the Markov partition satisfies the following additional property, the partition is said to be *expanding.*

(v) There exists $\lambda > 1$ such that for any $x, y \in I_i$ for some i, $d(f(x), f(y)) \geq \lambda d(x, y)$

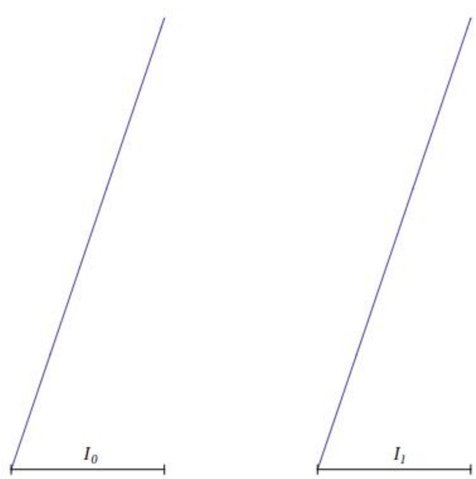

FIGURE 1. The map $L_3 : I_0 \cup I_1 \to [0, 1]$

A good picture to keep in mind for this is the discussion of last section: we had two intervals, which were in fact disjoint, and a map L_3 which brought their union to the entire interval $[0,1]$, as in Figure 1.

DEFINITION 5.2.2. Let $(f, \mathcal{P})$ be a map with a Markov Partition for some $f : X \to [0,1]$. Define $\Lambda = \Lambda_{(f,\mathcal{P})} = \Lambda(f)$ to be:

$$\Lambda = \{x \in X : f^n(x) \in X \ \forall n \in \mathbb{N}_0\}$$

EXERCISE 5.2.1. Show that Λ is a closed (and hence compact) subset of X, and that f always extends to a dynamical system $f : \Lambda \to \Lambda$.

THEOREM 5.2.3. *Let $f : X \to [0,1]$ be a continuous map with a full expanding Markov partition $\mathcal{P} = \{I_0, \ldots, I_{d-1}\}$. Then $\Lambda = \Lambda(f) \neq \emptyset$, and $f : \Lambda \to \Lambda$ is a dynamical system. Then there exists a surjection $h : \Sigma_d^+ \to \Lambda$ such that:*

(1) $h \circ \sigma = f \circ h$ (That is, h is a topological semiconjugacy)
(2) $h(\mathbf{x}) = y$ if and only if $f^{n-1}(y) \in I_{x_n}$ for every $n \geq 1$

PROOF. We first prove $\Lambda \neq \emptyset$. We know that for every i, j, $f(I_i) \supset I_j$. This implies that $f(I_i) \supset I_i$, so f has a fixed point in I_i by the intermediate value theorem. We now construct h. Denote $f|_{I_i}$ by f_i, so that $f_i : I_i \to f(I_i) \supset X$. Furthermore, since f_i is injective, we can define $g_i : f(I_i) \to I_i$ by letting $g_i = f_i^{-1}$. Let us see how we will define $h(\mathbf{0})$, where $\mathbf{0} = (0,0,0,0,0,\ldots)$. Then if $y_0 = h(\mathbf{0})$, we seek a point such that $f^n(y_0) \in I_0$ for every $n \geq 0$. That happens if and only if $y \in g_0^n(I_0)$ for every n. Then let $G_{\mathbf{0}}^n = g_0 \circ g_0 \circ \ldots \circ g_0 = g_0{}^n$. Thus our point y must lie in:

$$\bigcap_{n \geq 1} G_{\mathbf{0}}^{n-1}(I_0)$$

If we can show that the above set contains exactly one point, call it y, we can define $h(\mathbf{0}) = y$. We will prove that the sets above contain only one point in the general setting.

Equipped with these notions, let us move to the general case. Let $\mathbf{x} = (x_1, x_2, \ldots)$ be an arbitrary sequence. Define $G_{\mathbf{x}}^n = g_{x_1} \circ g_{x_2} \circ \ldots \circ g_{x_n}$. What happens if $y \in G_{\mathbf{x}}^{n-1}(I_{x_n})$? We claim that $f^{k-1}(y) \in I_{x_k}$ for $k = 1, \ldots, n$. Indeed, when $k = 1$, then y is in the image of g_{x_1}. That is, $y \in I_{x_1}$. If $k = 2$, then $f(y) = f \circ g_{x_1} \circ \ldots \circ g_{x_{n-1}}(I_{x_n}) = g_{x_2} \circ \ldots \circ g_{x_{n-1}}(I_{x_n})$. So $f(y) \in I_{x_2}$. Continuing by induction, we can see this goes up to $k = n$, when we get $f^{n-1}(y) \in I_{x_n}$. See Figure 2, as it applies to $L_3 : [0,1/3] \cup [2/3,1] \to [0,1]$.

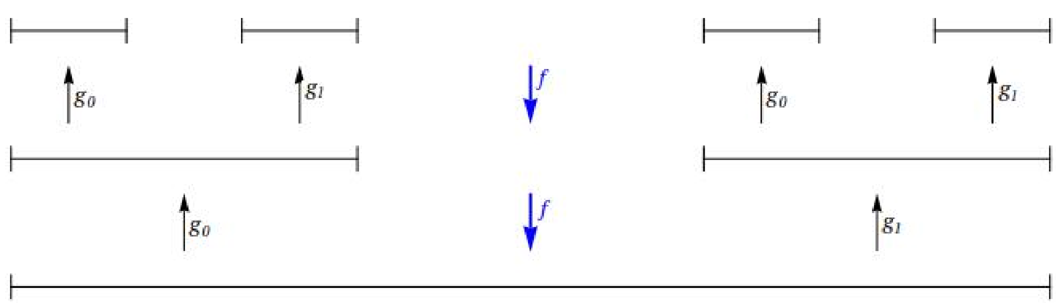

FIGURE 2. Iterating the maps g_i

Then we wish to consider:

$$\bigcap_{n\geq 1} G_{\mathbf{x}}^{n-1}(I_{x_n})$$

We wish to show that this set is nonempty and that the diameter of the sets $G_{\mathbf{x}}^{n-1}(I_{x_n}) \to 0$ (which will imply that the intersection is a single point. Indeed, we will show that the diameter approahes 0 exponentially fast! Let us first show that the intersection is nonempty. We claim that:

$$G_{\mathbf{x}}^{n-1}(I_{x_n}) \supset G_{\mathbf{x}}^{n}(I_{x_{n+1}})$$

This is showing that $g_{x_0} \circ \ldots \circ g_{x_{n-1}}(I_{x_n}) \supset g_{x_0} \circ \ldots \circ g_{x_n}(I_{x_{n+1}})$. But we know that $I_{x_n} \supset g_{x_n}(I_{x_{n+1}})$ by the full markov property! Hence we get a nested sequence of compact sets in X, whose intersection we know to be nonempty.

Let us show that their diameters tend to 0. Here we will use the expanding property of the partition. We claim that if $x, y \in X$, then $d(g_i(x), g_i(y)) < \lambda^{-1} d(x, y)$. To see this, note that if $x, y \in I_i$, then $d(f(x), f(y)) > \lambda d(x, y)$. But since g_i is f's local inverse on the set I_i, there exist $x' = f(x)$ and $y' = f(y)$ such that $x = g_i(x')$ and $y = g_i(y')$. Hence $d(g_i(x'), g_i(y')) < \lambda^{-1} d(x', y')$. Since f has the full markov property, every x' and y' have preimages, so this equality is satisfied everywhere.

This implies that $\operatorname{diam}(G_{\mathbf{x}}^n(I_{x_n})) \leq \lambda^{-n} \max\{\operatorname{diam}(I_i)\}$ by induction. This obviously tends to 0, so the intersection contains at most one point. To see continuity, let $\varepsilon > 0$. Choose n such that $\lambda^{-n} D < \varepsilon$ where $D = \max\{\operatorname{diam}(I_i)\}$. Then let $\delta = \beta^n(d-1)$, where β is the number chosen for the metric d_β. Then if $d_\beta(\mathbf{x}, \mathbf{y}) < \delta$, $x_k = y_k$ for $k = 0, \ldots, n$ by definition of the metric d_β. But this implies that $h(\mathbf{x}), h(\mathbf{y}) \in G_{\mathbf{x}}^{n-1}(I_{x_n})$ by construciton. Since we have shown that the diameter of set is less than $\lambda^{-n} D < \varepsilon$, we conclude that $d(h(\mathbf{x}), h(\mathbf{y})) < \varepsilon$. That is, h is continuous. □

Exercise 5.2.2. Show that $\Lambda(f)$ is the image of h, and that $f \circ h = h \circ \sigma$

Note that we have not used the fact that any two interval I_i and I_j intersect at most on their boundaries. We can use this property to detect when h is injective. Suppose that $h(\mathbf{x}) = h(\mathbf{y}) = y$. Then $f^n(y) \in I_{x_n} \cap I_{y_n}$ for every $n \geq 0$. Suppose that $\mathbf{x} \neq \mathbf{y}$, and $N = N(\mathbf{x}, \mathbf{y})$. Then $f^N(h(\mathbf{x})), f^N(h(\mathbf{y})) \in I_{x_N} \cap I_{y_N}$. Let $F = \{x : x \in I_j \cap I_i \text{ for some } , i, j\}$, and $C = \cup_{n\geq 0} f^{-n}(F)$. Then each points of $X \setminus C$ have a *unique* preimage on d.

5.3. Classification of Expanding Maps

The last section presented a theorem about continuous maps $f : X \to [0,1]$, where $X \subset [0,1]$ is a finite union of closed intervals, and the corresponding notion of Markov partitions (Definition 5.2.1 and Theorem 5.2.3). In fact, we can extend this notion to the following class of maps:

Definition 5.3.1. Let X be a metric space, and $f : X \to X$ a dynamical system. f is *expanding* if there exists some $\varepsilon > 0$ and $\lambda > 1$ such that if $d(x, y) < \varepsilon$, then $d(f(x), f(y)) \geq \lambda d(x, y)$

The inclusion of the ε may seem mysterious, but let us see its necessity from the viewpoint of the simple maps L_d. We call them expanding because they take small intervals and map them into intervals of length d, essentially making them expanding with constant $\lambda = d$. But even on L_2, $0 = d([0],[0]) = d(f([0]), f([1/2])) < d(0, 1/2) = 1/2$!

First, let's see the complications that prevent us from using Theorem 5.2.3 directly, even for the maps L_d. In these maps, we wish to use the partition $\mathbb{R}/\mathbb{Z} = I_0 \cup \cdots \cup I_{d-1}$, where $I_i = [i/d, (i+1)/d]$. In the definition of full Markov partition we require that $f|_{I_i}$ is a homeomorphism. Unfortunately, $f(i/d) = f((i+1)/d)$, so f is not a homeomorphism. One idea would be let $g : [0,1] \to [0,1]$ be defined by $g(x) = f([x])$. Unfortunately, again, we have a problem, because at the point $1/d$, we don't know whether to define $g(1/d) = 0$ or 1. So we have to think of a new scheme. First, we need a technical proposition:

PROPOSITION 5.3.2. *A map $f : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ is expanding if and only if there exists some $\lambda > 1$ such that for every lift $\tilde{f}$ of f and $x, y \in \mathbb{R}$: $d(\tilde{f}(x), \tilde{f}(y)) \geq \lambda d(x, y)$. Furthermore, $|\deg(f)| > 1$, and every lift of f is injective*

PROOF. Let f be an expanding map, and $\tilde{f}$ be a lift of f. We claim that f either strictly increasing or decreasing. Suppose otherwise, and that there exists points $x_- < x_0 < x_+ \in \mathbb{R}$such that $x_- \leq x_0$ and $x_+ \leq x_0$, and $\tilde{f}(x_0)$ achieves its maximum at x_0 in $B_\varepsilon(x_0)$. Without loss of generality, by the intermediate value theorem, we can assume that x_- and x_+ are arbitrarily close to x_0 and that $f(x_-) = f(x_0)$. But then, since x_- and x_+ are close to x_0, $[x_-]$, $[x_0]$ and $[x_+]$ are all distinct points of $\mathbb{R}/\mathbb{Z}$, and $d([x_-], [x_+])$ can be made less than ε, but $d(f([x_-]), f([x_+])) = 0$. This implies that f is not expanding, hence every lift is strictly increasing or decreasing (and hence injective).

Suppose that $\tilde{f}$ is strictly increasing. Then let $\varepsilon > 0$ be as in the definition of expanding, and without loss of generality, assume $\varepsilon < 1/2$. If $x < y \in \mathbb{R}$, choose $N > (y - x)/\varepsilon$. Then the points $z_k = (kx + (N - k)y)/N$ satisfy $d(z_k, z_{k+1}) < \varepsilon$, and hence:

$$d(\tilde{f}(x), \tilde{f}(y)) = \sum_{k=0}^{N} \tilde{f}(z_{k+1}) - \tilde{f}(z_k) = \sum_{k=0}^{N} d(f([z_{k+1}]), f([z_k])) \geq \sum_{k=0}^{N} d([z_{k+1}], [z_k]) = d(x, y)$$

Now suppose that f has a lift $\tilde{f}$ satisfying the desired properties. Let $\mathcal{U} = \left\{B_{1/4}([x]) : [x] \in \mathbb{R}/\mathbb{Z}\right\}$ be the open cover of $\mathbb{R}/\mathbb{Z}$ by balls of radius $1/4$. Let $\mathcal{V} = \tilde{f}^{-1}(\mathcal{U}) = \left\{f^{-1}(B_{1/4}(x)) : x \in \mathbb{R}/\mathbb{Z}\right\}$. Then $\mathcal{V}$ is an open cover of $\mathbb{R}/\mathbb{Z}$ and has a Lesbesgue number ε, which we can without loss of generality can assume is less than $1/2$. Then if $d([x], [y]) < \varepsilon$, $d(x, y) < \varepsilon$ for suitable choices of x and y. Furthermore, $d([f(x)], [f(y)]) < 1/4$ by construction, so $d(\tilde{f}(x), \tilde{f}(y)) = d(f([x]), f([y]))$. So $d(f([x]), f([y])) = d(\tilde{f}(x), \tilde{f}(y)) \geq \lambda d(x, y) = \lambda d([x], [y])$. So f is expanding.

One now easily computes that $|\deg(f)| = \left|\tilde{f}(1) - \tilde{f}(0)\right| \geq \lambda$.

□

THEOREM 5.3.3. *Let f be an expanding map of degree d. Then there exists a topological conjugacy $h : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ such that $f \circ h = h \circ L_d$*

Before proving the Theorem, let's observe a few immediate corollaries:

COROLLARY 5.3.4. *If f is an expanding map of degree d, there is an "almost one-to-one" topological semiconjugacy $h : \Sigma_d^+ \to \mathbb{R}/\mathbb{Z}$ such that $h \circ \sigma = f \circ h$*

PROOF. This follows from the fact that we have a canonical projection $h_0 : \Sigma_d^+ \to \mathbb{R}/\mathbb{Z}$ such that $h_0 \circ \sigma = L_d \circ h_0$. In this case, if h_1 is the conjugacy such that $h_1 \circ L_d = f \circ h_1$, then set $h = h_1 \circ h_0$. Then $h \circ \sigma = h_1 \circ h_0 \circ \sigma = h_1 \circ L_d \circ h_0 = f \circ h_1 \circ h_0 = f \circ h$. □

COROLLARY 5.3.5. *If f and g are expanding maps of the same degree, then they are topologically conjugate*

PROOF. Let h_f and h_g be the conjugacies such that $f \circ h_f = h_f \circ L_d$ and $g \circ h_g = h_g \circ L_d$. Then let $h = h_f \circ h_g^{-1}$, so that $h \circ g = h_f \circ h_g^{-1} \circ g = h_f \circ L_d \circ h_g^{-1} = f \circ h_f \circ h_g^{-1} = f \circ h$. □

There is one more important notion that should be mentioned. We need a definition:

DEFINITION 5.3.6. Let $\mathcal{C}(X, Y)$ be the set of continuous functions from X to Y. Then define $C^1(\mathbb{R}/\mathbb{Z}, \mathbb{R}/\mathbb{Z}) = \left\{f \in \mathcal{C}(\mathbb{R}/\mathbb{Z}, \mathbb{R}/\mathbb{Z}) : \exists \tilde{f} \text{ which is continuously differentiable}\right\}$. Define a metric on $C^1(\mathbb{R}/\mathbb{Z}, \mathbb{R}/\mathbb{Z})$ by setting:

$$d_{C^1}(f, g) = \max\left\{d(f([x]), g([x])), \left|\tilde{f}'(x) - g'(x)\right| : x \in \mathbb{R}\right\}$$

Furthermore, let $\mathcal{E} \subset \mathcal{C}(\mathbb{R}/\mathbb{Z}, \mathbb{R}/\mathbb{Z})$ denote the set of expanding maps.

This space makes the property of expanding have a very nice interpretation:

PROPOSITION 5.3.7. *$f \in \mathcal{E} \cap C^1(\mathbb{R}/\mathbb{Z}, \mathbb{R}/\mathbb{Z})$ if and only if there exists a lift $\tilde{f}$ of f such that $\left|\tilde{f}'(x)\right| \geq \lambda$ for every $x \in \mathbb{R}$ and some $\lambda > 1$*

PROOF. First, suppose that the derivative is bounded from below. Then:

$$d(\tilde{f}(x), \tilde{f}(y)) = \left|\tilde{f}(x) - \tilde{f}(y)\right| = \left|\int_x^y f'(t)\,dt\right| = \int_x^y |f'(t)|\,dt \geq \int_x^y \lambda\,dt = \lambda d(x,y)$$

We make use of the fact, that since $|f'(x)| \geq \lambda$ everywhere, it must be either always positive or always negative, and hence we can pull the absolute value bars inside the integral.

Now suppose that f is expanding. Then for $x \in \mathbb{R}$:

$$|f'(x)| = \left|\lim_{t\to 0} \frac{f(x+t) - f(x)}{t}\right| \geq \lim_{t\to 0} \frac{\lambda\,|x+t-x|}{t} = \lim_{t\to 0} \lambda = \lambda$$

□

We get a very special property of these maps, beyond the discussions of this course, called *structural stability*. We present it as a Corollary, but will not define the general notion:

COROLLARY 5.3.8. *Let $f \in \mathcal{E} \cap C^1(\mathbb{R}/\mathbb{Z}, \mathbb{R}/\mathbb{Z})$. Then there exists some $\varepsilon > 0$ such that if $d_{C^1}(g,f) < \varepsilon$, then f and g are topologically conjugate*

PROOF. Let λ be the constant for f as in the expanding map definition. Choose $\varepsilon < \min \lambda - 1, 1/2$. Then if $d_{C^1}(g,f) < \varepsilon$, $|g'(x)| = |g'(x) - f'(x) + f'(x)| \geq |f'(x)| - |g'(x) - f'(x)| \geq \lambda - \varepsilon > 1$. Hence, g is expanding. Furthermore:

$$|\deg(f) - \deg(g)| \leq \left|\tilde{f}(1) - \tilde{g}(1)\right| + \left|\tilde{f}(0) - \tilde{g}(0)\right| < 1$$

Hence $\deg(f) = \deg(g)$ since they must both be integers. So by the Theorem, f and g are topologically conjugate. □

After stalling sufficiently, let us finally provide an exposition of the proof. It is rather technical, but keep in mind the proof of Theorem 5.2.3, as the arguments are parallel.

PROOF OF THEOREM 5.3.3. The proof idea is similar to that of the proof of Theorem 5.2.3. We wish to use the infinite intersection of sets to define our map. Let us first analyze the linear expanding map L_d. For each finite sequence $\mathbf{x} = (x_1, \ldots, x_n)$, define $|\mathbf{x}| = n$, and $\mathcal{F}(n) = \{\mathbf{x} : |\mathbf{x}| = n\}$ be the set of finite sequences of length n.

For each $\mathbf{x} \in \mathcal{F}(n)$, let $\Delta_{\mathbf{x}} = [\mathbf{x}^-, \mathbf{x}^+]$, where $\mathbf{x}^+ = \sum_{k=1}^n x_k d^{-k}$ and $\mathbf{x}^+ = \mathbf{x}^- + d^{-n}$. We allow the empty word $\mathbf{x} = \emptyset$ and set $\Delta_\emptyset = [0,1] = \mathbb{R}/\mathbb{Z}$, and define the function $\sigma : \mathcal{F}(n) \to \mathcal{F}(n-1)$ by $\sigma(x_1, \ldots, x_n) = (x_2, \ldots, x_n)$. One can easily compute directly that $L_d(\Delta_{\mathbf{x}}) = \Delta_{\sigma(\mathbf{x})}$.

Now if $t \in \mathbb{R}/\mathbb{Z}$ has d-ary expansion $t = \sum_{k=0}^\infty t_k d^{-k}$, then t is the unique element of $\bigcap_{n\geq 1} \Delta_{\tau_n(\mathbf{t})}$, where $\tau_n : \Sigma_d^+ \to \mathcal{F}(n)$ is just the function which takes an infinite sequence to its first n entries. We also know that a number t has more than one d-ary representation $\mathbf{t}$ and $\mathbf{t}'$ if and only if $t_N = t'_N + 1$ and $t_n = 0$, $t'_n = d - 1$ for $n > N$.

We seek similar structures for the map f. We assume that $d > 0$, the negative case has parallel arguments. Let $\tilde{f}$ be a lift of f. Then if $F(x) = \tilde{f}(x) - x$, $F(0) = \tilde{f}(0)$ and $F(1) = \tilde{f}(1) - 1 = \tilde{f}(0) + (d-1)$. Since $d > 1$, $F([0,1]) \supset [F(0), F(0) + (d-1)]$. Such an interval must contain an integer, so by changing the lift of f by this integer, we can assume that F has a fixed point in $[0,1]$, call it c_0. Consider $f^{-1}([c_0])$. These point can be identified with points of $[0,1)$ which are mapped to $c_0 + \mathbb{Z}$. Since $\tilde{f}([0,1)) = [\tilde{f}(0), \tilde{f}(0) + d)$, and the map is injective, there are exactly d preimages, write them in increasing order as $c_0 < c_1 < \cdots < c_{d-1}$. Let $\Gamma_{(i)} = [c_i, c_{i+1}]$ for $i = 0, \ldots, d-2$ and $\Gamma_{(d-1)} = [c_{d-1}, c_0 + 1]$.

Without loss of generality, we claim that we may assume that $c_0 = 0$. Indeed, take the map $h_0 : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ to be $h([x]) = [x - c_0]$. Then the map $h_0 \circ f \circ h_0^{-1}$ has $h_0(f(h_0^{-1}(0))) = h_0(f(c_0)) = h_0(c_0) = 0$. The map $h_0 \circ f \circ h_0^{-1}$ is again an expanding map since h_0 is a bijective isometry. Hence, if we can find a conjugating map of $h_0 \circ f \circ h_0^{-1}$ say h_1, we can find one for f by setting $h = h_1 \circ h_0$.

We now assume that $c_0 = 0$ so that $f(0) = 0$. Choose the lift $\tilde{f}$ of f such that $\tilde{f}(0) = 0$. Then $\tilde{f}(c_i) = i$ by construction, so $\tilde{f}(\Gamma_{(i)}) = [i, i+1]$. This implies that $f(\Gamma_{(i)}) = \mathbb{R}/\mathbb{Z}$ and the map is injective with the exception of the boundary terms c_i and c_{i+1}. Thus, $f^{-1}\left(\Gamma_{(i)}\right)$ will be a disjoint union of intervals $\Gamma_{(i,j)}$ where j ranges from 0 to $d-1$ and $\Gamma_{(i,j)} = \Gamma_{(i)} \cap f^{-1}\left(\Gamma_{(j)}\right)$. Proceeding by induction, for each $\mathbf{x} \in \mathcal{F}(n)$, we can define an interval $\Gamma_{\mathbf{x}}$ such that $f(\Gamma_{\mathbf{x}}) \subset \Gamma_{\sigma(\mathbf{x})}$. Indeed, if $\Gamma_{\sigma(\mathbf{x})}$ is well-defined, set $\Gamma_{\mathbf{x}} = f^{-1}(\Gamma_{\sigma(\mathbf{x})}) \cap \Gamma_{(x_1)}$.

Let $D = \max\left\{\operatorname{diam}\Gamma_{(j)} : j = 0, \dots, d-1\right\}$. We claim that $\operatorname{diam}\Gamma_{\mathbf{x}} \leq \lambda^{|x|}D$. For $|x| = 1$, this is exactly the definition. Now if $\operatorname{diam}\Gamma_{\sigma(\mathbf{x})} < \lambda^{1-|x|}D$, and $s, t \in \Gamma_{\mathbf{x}}$, then $f(s)$ and $f(t)$ are in $\Gamma_{\sigma(\mathbf{x})}$ and $d(f(s), f(t)) < \lambda^{1-|x|}D$. Then if $\tilde{f}$ is a lift of f, $d(\tilde{f}(s), \tilde{f}(t)) < \lambda^{1-|x|}D$, where $\tilde{f}(s)$ and $\tilde{f}(t)$ may not lie in the same lift of $\Gamma_{\sigma(\mathbf{x})}$, but still project to $f(s)$ and $f(t)$ on $\mathbb{R}/\mathbb{Z}$. This implies that $d(s,t) < \lambda^{-1} d(\tilde{f}(s), \tilde{f}(t)) < \lambda^{-|x|}D$ on $\mathbb{R}$. Hence $d(s,t) < \lambda^{-|x|}D$ on $\mathbb{R}/\mathbb{Z}$ as well.

So we have a collection of closed intervals $\Gamma_{\mathbf{x}}$, for $\mathbf{x} \in \mathcal{F}(n)$. Let $\mathbf{x}$ now represent an infinite sequence. That is, $\mathbf{x} \in \Sigma_d^+$. We claim that $\Gamma_{\tau_n(\mathbf{x})} \supset \Gamma_{\tau_{n+1}(\mathbf{x})}$. Note that:

$$\Gamma_{\tau_n(\mathbf{x})} = \Gamma_{(x_1)} \cap \Gamma_{\sigma(\tau_n(\mathbf{x}))} = \Gamma_{(x_1)} \cap f^{-1}\left(\Gamma_{(x_2)}\right) \cap \Gamma_{\sigma^2(\tau_n(\mathbf{x}))} = \cdots = \Gamma_{(x_1)} \cap f^{-1}\left(\Gamma_{(x_2)}\right) \cap \cdots \cap f^{-n+1}\left(\Gamma_{(x_n)}\right)$$

From this formula, the inclusion is obvious. Now the path is clear: For each $\mathbf{x} \in \Sigma_d^+$, let $H(\mathbf{x}) = \bigcap_{n\geq 1} \Gamma_{\tau_n(\mathbf{x})}$. Since these are nested, compact sets, whose diameter tends to 0, they contain exactly one point, so the function is well defined. Suppose now that $H(\mathbf{x}) = H(\mathbf{y})$. Then $\bigcap_{n\geq 1} \Gamma_{\tau_n(\mathbf{x})} = \bigcap_{n\geq 1} \Gamma_{\tau_n(\mathbf{y})}$. Suppose $\mathbf{x} \neq \mathbf{y}$, and let N denote the first time that these sequences differ. Then $f^{N-1}(\Gamma_{\tau_N(\mathbf{x})}) = \Gamma_{(x_N)}$, so $\Gamma_{(x_N)} \cap \Gamma_{(y_N)} \neq \emptyset$. But if $x_N \neq y_N$, and $x_{N-1} = y_{N-1}$, this implies that $x_N = y_N + 1$ (or the other way around), and $f^{N-1}(H(\mathbf{x})) = \Gamma_{(x_N)} \cap \Gamma_{(y_N)}$. In this case, $f^N(H(\mathbf{x})) = f^N(H(\mathbf{y}))) = 0$, so $H(\sigma^N(\mathbf{x})) = H(\sigma^N(\mathbf{y})) = 0$.

We claim that the only sets $\Gamma_{\mathbf{x}}$ containing 0 are those which contain all 0's or all $(d-1)$'s. Note that if $t \in \Gamma_{\mathbf{x}}$, then $f^n(t) \in \Gamma_{(x_{n+1})}$. Since 0 is fixed, $0 \in \Gamma_{(x_n)}$ for every $n \geq 1$. This implies $x_n = 0$ or $d-1$ for every n. Suppose that $x_n = 0$ and $x_{n+1} = d-1$ (the other case follows similarly). We claim then that $0 \notin \Gamma_{\mathbf{x}}$. It suffices to show it for the case when $|x| = 2$ by the nesting properties of the $\Gamma_{\mathbf{x}}$ and the fact that 0 is fixed. That is, we wish to show that $\Gamma_{(0)} \cap f^{-1}(\Gamma_{(d-1)})$ does not contain 0. Indeed, f restricted to this set is a homeomorphism onto $\Gamma_{(d-1)}$, of which 0 is in. Indeed, $f^{-1}(\Gamma_{(d-1)}) \cap \Gamma_{(0)} = [f^{-1}(c_{d-2}), c_1]$. 0 is clearly not in this set, so any 0 cannot be followed by a $d-1$. Similarly, $d-1$ cannot be followed by a 0. Thus the only way for $H(\mathbf{x}) = H(\mathbf{y})$ nontrivially is the same as that of the map for L_d.

We can now define our function h_1 by first expressing our point $t \in \mathbb{R}/\mathbb{Z}$ as $t = \sum_{k\geq 1} t_k d^{-k}$, and then letting $h_1(t)$ be the point $H(\mathbf{t}) \in \mathbb{R}/\mathbb{Z}$. The discussion above shows that $h_1(t)$ is independent of the choice of d-ary expansion. Also, let $S(\mathbf{t}) = \sum_{k\geq 1} t_k d^{-k}$, so that $h_1 \circ H = S$. By construction, $H \circ \sigma = f \circ H$ and $S \circ \sigma = L_d \circ S$. Then, if $H(\mathbf{x}) = x$:

$$h_1 \circ f(x) = h_1 \circ f \circ H(\mathbf{x}) = h_1 \circ H \circ \sigma(\mathbf{x}) = S \circ \sigma(\mathbf{x}) = L_d \circ S(\mathbf{x}) = L_d \circ h_1 \circ H(\mathbf{x}) = L_d \circ h_1(x)$$

The map h_1 is a homeomorphism if it is continuous, as it is bijective and $\mathbb{R}/\mathbb{Z}$ is compact. It takes the closed sets $\Gamma_{\mathbf{x}}$ to the closed sets $\Delta_{\mathbf{x}}$. Thus it takes maps their interiors to interiors. Furthermore, if t is on the boundary of one of these sets, then $\Gamma_{\mathbf{x}} \cap \Gamma_{\mathbf{x}_+} = t$, where $\mathbf{x}_+ = (x_1, \dots, x_n + 1)$. Since in this case, $h_1(\Gamma_{\mathbf{x}} \cup \Gamma_{\mathbf{x}_+}) = \Delta_{\mathbf{x}} \cup \Delta_{\mathbf{x}_+}$ and hence their interiors to the interiors, we can also take arbitrarily small neighborhoods of t to open sets. Since any open set of $\mathbb{R}/\mathbb{Z}$ can be expressed as a union of sets of this form, the map h_1 takes open sets to open sets. Hence it is continuous.

$\square$

5.4. Maps of Large Degree

This is not the end of our story for circle maps! Expanding maps have very nice behvaior, but not every map of degree > 1 is expanding. We need a (necessarily more complicated) Theorem to describe the general setting:

THEOREM 5.4.1. *If $f : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ has $|\deg(f)| \geq 2$, then there is a semiconjugacy $h : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ of degree 1 such that $h \circ f = L_d \circ h$. Furthermore, such a map h is unique up to translation. That is, if $\bar{h}$ is another such semiconjugacy, then $\bar{h}(x) - h(x) \equiv y$ for some $y \in \mathbb{R}/\mathbb{Z}$*

Let us employ a different proof strategy. Note that the equation $h \circ f = L_d \circ h$, if satisfied, should lift to an equation $\tilde{h} \circ \tilde{f} = \widetilde{L_d} \circ \tilde{h}$. Since we want h to have degeree 1, we should be able to write it as $\tilde{h}(x) = x + u(x)$ for some 1-periodic $u(x)$. We can also write $\tilde{f}(x) = dx + \varphi(x)$ Then our equation becomes:

$$\begin{aligned}
\tilde{h} \circ \tilde{f}(x) &= \widetilde{L_d} \circ \tilde{h}(x) \\
\tilde{f}(x) + u(\tilde{f}(x)) &= d\tilde{h}(x) \\
dx + \varphi(x) + u(\tilde{f}(x)) &= dx + du(x) \\
\varphi(x) + u(\tilde{f}(x)) &= du(x) \\
u(x) - \frac{1}{d}u(\tilde{f}(x)) &= \frac{\varphi(x)}{d}
\end{aligned}$$

Now we need to enter some elementary functional analysis. Let $V = \{u : \mathbb{R} \to \mathbb{R} \text{ such that } u \text{ is 1-periodic }\}$. Then V is a vector space, and $L : V \to V$ defined by $L(u)(x) = \frac{1}{d}u \circ \tilde{f}$. Then if $w = \frac{\varphi(x)}{d}$. Writing this in linear algebraic terms, we wish to solve:

$$(\mathrm{id} - L)u = w$$

This, if we can make sense of $(\mathrm{id} - L)^{-1}$, we can solve the problem. We recognize that in the reals, we can write: $(1-x)^{-1} = \frac{1}{1-x} = 1 + x + x^2 + \dots$ whenever $|x| < 1$. This motivates us to define:

$$u(x) = \sum_{n \geq 0} L^n w = \sum_{n \geq 0} \frac{\varphi \circ \tilde{f}^n(x)}{d^{n+1}}$$

How can we guarantee that such limits exist and the function u is well-defined? The Weierstrass M-criterion will suffice:

THEOREM 5.4.2 (Weiertrass M-Criterion). *Let $u_n : \mathbb{R} \to \mathbb{R}$ be 1-periodic and satisfy $|u_n(x)| \leq a_n$ and $\sum_{n \geq 0} a_n \leq M < \infty$, then $u(x) = \sum_{n \geq 1} u_n(x)$ is continuous*

We will omit the proof of this Theorem, but let us see that u satisfies our desired properties by computing directly that:

$$u(x) - \frac{1}{d}u \circ \tilde{f}(x) = \sum_{n \geq 0} \frac{\varphi \circ \tilde{f}^n(x)}{d^{n+1}} - \frac{1}{d}\sum_{n \geq 0} \frac{\varphi \circ \tilde{f}^{n+1}(x)}{d^{n+1}} = \sum_{n \geq 0} \frac{\varphi \circ \tilde{f}^n(x)}{d^{n+1}} - \sum_{n \geq 1} \frac{\varphi \circ \tilde{f}^n(x)}{d^{n+1}} = \frac{\varphi(x)}{d}$$

Let us move now to uniqueness. Suppose that $g : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ also satisfies that $g \circ f = L_d \circ h'$, and $\deg(g) = 1$. Again, we lift our equality using lifts of f and g, so that:

$$\tilde{g} \circ \tilde{f} = d\tilde{g} + n$$

for some fixed $n \in \mathbb{Z}$. Our solution h satisfied this, as well, so subtracting:

$$(\tilde{g} - \tilde{h}) \circ \tilde{f} = d(\tilde{g} - \tilde{h}) + n$$

Since g and h have degree 1, $\psi = \tilde{g} - \tilde{h}$ is 1-periodic. In particular:

$$\psi \circ \tilde{f} = d\psi + n$$

Suppose first that ψ is constant, ie $\psi \equiv a$. Then our equation becomes $a = da + n$, so $a = n/(1-d)$. Thus, if we define $\psi_a = \psi - a$, we get that $\psi_a \circ \tilde{f} = d\psi_a$ by computing directly. Let $M = \max_{\mathbb{R}} |\psi_a|$ (this maximum must exist since ψ_a is 1-periodic). Let x_0 satisfy, $|\psi_a(x_0)| = M$. But then $\left|\psi_a(\tilde{f}(x_0))\right| = |d\psi_a(x_0)| = |d|\, M$. So $M = 0$, and $\psi_a \equiv 0$. All solutions must then take the form:

$$g = h + \frac{k}{d-1}$$

with $k = 0, \ldots, d-2$.

We can indeed use this method to find an alternative proof of Theorem 5.3.3.

Alternate proof of Theorem 5.3.3. In light of Theorem 5.4.1, we only need to verify that the semiconjugacy obtained is injective, as if f is expanding, it has $|\deg(f)| \geq 2$ (Proposition 5.3.2). Choose a lift $\tilde{h}$ of the semiconjugacy. By choosing a nice lift, we can assume that $\tilde{h} \circ \tilde{f} = d\tilde{h}$. Then $\tilde{h}(\tilde{f}^n(x)) = d^n\tilde{h}(x) = d^n\tilde{h}(y) = \tilde{h}(\tilde{f}^n(y))$ for every $n \in \mathbb{N}_0$. Write $h(x) = x + u(x)$ for some 1-periodic u. So $\tilde{f}^n(x) + u(\tilde{f}^n(x)) = \tilde{f}^n(y) + u(\tilde{f}^n(y))$, and:

$$d(\tilde{f}^n(x), \tilde{f}^n(y)) = \left|\tilde{f}^n(x) - \tilde{f}^n(y)\right| = \left|u(\tilde{f}^n(x)) - u(\tilde{f}^n(y))\right|$$

But since u is 1-periodic, it is bounded, so $d(\tilde{f}^n(x), \tilde{f}^n(y)) \leq C$ for a fixed constant C. Since $\tilde{f}$ is expanding, this means $\lambda^n d(x,y) \leq d(\tilde{f}^n(x), \tilde{f}^n(y)) \leq C$, so $d(x,y) \leq C\lambda^{-n}$ for every $n \geq 0$. This implies that $d(x,y) = 0$ and hence h is injective.

$\square$

CHAPTER 6

Homeomorphisms and Rotation Number

In this section, we will study homeomorphisms of the circle. We have seen, that if $h : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ is a homeomorphism, it must have $\deg(h) = \pm 1$. The main example we have seen are the rotations $R_\alpha(x) = x + \alpha$. Another example with simpler dynamics is the North-South Pole map:

$$h([x]) = \frac{1}{2}\left(1 - \cos(\pi x)\right)$$

Here, we identify $[x]$ with its representative in $[0, 1)$. We plot this map in two ways. First, using the interval $[0, 1]$ with endpoints identified, we can visualize it as a circle map in Figure 1. We see that the map has two fixed points, and that every other point moves away from the fixed point 0 (the north pole) and towards the fixed point $1/2$ (the south pole). We see exactly this in the first few iterates of the map seen on the circle (the arrows represent the map f in Figure 2).

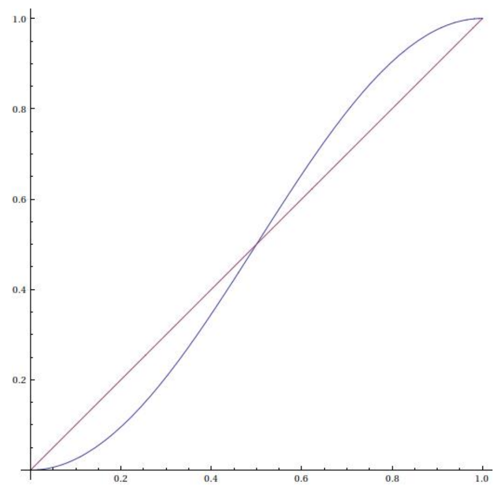

FIGURE 1. The Interval Picture of h

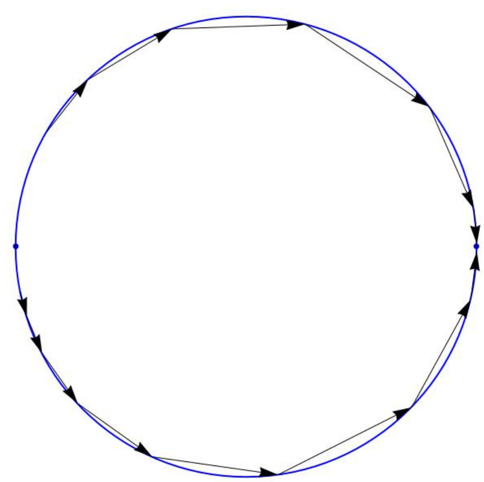

FIGURE 2. Viewing h on S^1

These two maps have very different behavior. In rotations, we "move" a fixed amount in the same direction around the circle. But for the North-South Pole map, the amount the point moves becomes very small... so small that we never leave a sector of the circle. In this sense, there is no rotation. We present the following definition, and its justification will soon follow:

DEFINITION 6.0.3 (Rotation Number). Let $f : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ be a degree one homeomorphism. Choose a lift $\tilde{f} : \mathbb{R} \to \mathbb{R}$ of f, and define:

$$\rho(f) = \lim_{n \to \infty} \frac{1}{n}\left(\tilde{f}^n(x) - x\right) \pmod 1$$

$\rho(f)$ is called the *rotation number* of f

Let us observe first that this deserves to be called "rotation number." Note that:

$$\frac{1}{n}\left(\tilde{f}^n(x) - x\right) = \frac{1}{n}\sum_{k=1}^{n} \tilde{f}^k(x) - \tilde{f}^{k-1}(x)$$

So the rotation number is indeed the average rotation for the point $x \in \mathbb{R}/\mathbb{Z}$. The number $\rho(f)$, to be well-defined, must satisfy the following properties:

(i) The limit in the definition exists
(ii) ρ is independent of the choice of lift $\tilde{f}$
(iii) ρ is independent of $x \in \mathbb{R}/\mathbb{Z}$

Parts (i) and (iii) are implied by:

PROPOSITION 6.0.4. *Let $g_n(x) = \frac{1}{n}\left(\tilde{f}^n(x) - x\right)$. Then $g_n(x)$ converges uniformly to some constant function*

First, let's show (ii) assuming (i) and (iii). Let F_1 and F_2 be two lifts of f, and ρ_1 and ρ_2 the rotation numbers obtained from such lifts. Then $F_2(x) = F_1(x) + c$ for some $c \in \mathbb{Z}$:

$$\rho_1(f) = \lim_{n\to\infty} \frac{1}{n}\left(F_1^n(x) - x\right) = \lim_{n\to\infty}\frac{1}{n}\left(F_2^n(x) + cn - x\right) = \rho_2(f) + c$$

Thus, since $\rho(f)$ is the projection of these real numbers into $\mathbb{R}/\mathbb{Z}$, $\rho_1 = \rho_2 \pmod 1$.

In the proof of Proposition 6.0.4, we will make use of the following Lemma (to appear in the homework set):

LEMMA 6.0.5. *Let $a_n \in \mathbb{R}$ be a sequence of real numbers, and suppose that there is some $L > 0$ such that $a_{m+n} \le a_m + a_n + L$. Then $\lim_{n\to\infty} \frac{a_n}{n}$ converges to some $r \in \mathbb{R} \cup \{-\infty\}$.*

PROOF OF PROPOSITION 6.0.4. Let $g : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ be a degree one homeomorphism, and $\tilde{g}$ a lift of g. Then $\tilde{g}$ is strictly increasing, so if $x < y < x+1$, then $\tilde{g}(x) < \tilde{g}(y) < \tilde{g}(x+1) = \tilde{g}(x)+1$, so $\tilde{g}(y) - \tilde{g}(x) \in [0,1]$. This implies that if $x, y \in [0,1]$, then $d(\tilde{g}(x), \tilde{g}(y)) \le 1$. Thus, if $u(x) = \tilde{g}(x) - x$:

$$d(u(x), u(y)) \le d(\tilde{g}(x), \tilde{g}(y)) + d(x,y) \le 2$$

An important aspect is that this inequality is independent of g. In particular, if we write $\tilde{f}^n(x) = x + u_n(x)$, we know that $d(u_n(x), u_n(y)) \le 2$, so $\frac{1}{n}\left|u_n(x) - u_n(y)\right| \le \frac{2}{n}$, so if the functions u_n converge pointwise, they converge to a constant function in the uniform topology.

Fix $x \in \mathbb{R}$. Let $a_n = u_n(x)$. Then:

$$\begin{aligned}
a_{m+n} &= u_{m+n}(x)\\
&= \tilde{f}^{m+n}(x) - x\\
&= \tilde{f}^{m+n}(x) - \tilde{f}^m(x) + \tilde{f}^m(x) - x\\
&= u_n(\tilde{f}^m(x)) + u_m(x)\\
&\le u_n(x) + u_m(x) + \left|u_n(\tilde{f}^m(x)) - u_n(x)\right|\\
&\le a_m + a_n + 2
\end{aligned}$$

Thus $\rho(f)$ exists, but may be $-\infty$. Note that $u_n(x) = \sum_{k=1}^{n} u_1(\tilde{f}^{k-1}(x))$. Since u_1 is 1-periodic, then $u_1 \ge R$ for some constant R. Thus $u_n \ge nR$ and $\frac{1}{n}u_n \ge R$, so the limit cannot be $-\infty$.

□

This proof demonstrated a common technique in dynamical systems which we call the *Cesàro Average*. If α_n is a sequence of real numbers, its Cesàro average is the number

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}\alpha_k$$

EXERCISE 6.0.1. Show that if a sequence converges, its Cesàro average is the same value. Find a sequence which does not converge but has a convergent Cesàro average

6.1. Rotation Number and Periodic Points

Throughout this section, we will use two notions of rotation number. Note that in the definition of the rotation number, we chose a lift $\tilde{f}$ of a homeomorphism f, and showed that we can associate a unique element of $\mathbb{R}$ to the lift $\tilde{f}$, $\rho_0(\tilde{f})$. Then we showed that this element was unique as an element of $\mathbb{R}/\mathbb{Z}$ (but not of $\mathbb{R}$). We will let $\rho_0(\tilde{f}) \in \mathbb{R}$ denote the rotation number of a lift, and $\rho(f) \in \mathbb{R}/\mathbb{Z}$ the rotation number of the homeomorphism f.

PROPOSITION 6.1.1. *If $f : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ is a degree one homeomorphism with a periodic point, then $\rho(f) \in \mathbb{Q}$*

PROOF. Let x be such that $f^n(x) = x$, so that $\tilde{f}^n(x) = x+k$. This implies that $\tilde{f}^{ln}(x) = x+lk$, so $\lim_{l\to\infty}\frac{1}{ln}\left(\tilde{f}^{ln}(x) - x\right) = k/n = \rho(f)$. □

In fact, the converse is true.

PROPOSITION 6.1.2. *Let $f : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ be a homeomorphism of degree 1. Then if $\rho(f) = p/q \in \mathbb{Q}$ (here p and q are coprime), f has a point of period q such that for some lift $\tilde{f}$ of f: $\tilde{f}^q(x) = x + p$. Furthermore, for any periodic point, $\tilde{f}^q(y) = y + p$*

The proof of the proposition will follow from the following Lemmas:

LEMMA 6.1.3. *Let g be a homeomorphism of degree 1 on $\mathbb{R}/\mathbb{Z}$, then $\rho(g) = 0$ if and only if there exists a lift $\tilde{g}$ of g such that $\tilde{g}$ has a fixed point.*

PROOF. Assume, for contradiction, that for every lift $\tilde{g}$ of g, $u(x) = \tilde{g}(x) - x \neq 0$. Then either $u > 0$ or $u < 0$ everywhere. Since u is 1-periodic, it takes its minimum and maximum, ie either $u(x) \geq c > 0$ or $u(x) \leq c < 0$. We will prove the case of $u(x) \leq c < 0$ (the other has a symmetric argument). Then $\rho_0(\tilde{g}) = \lim_{n\to\infty} u_n(x)/n$, where $u_n(x) = \tilde{g}^n(x) - x = \sum_{k=0}^{n-1} u(\tilde{g}^k(x)) \leq nc$. Thus, $\rho(\tilde{g}) \leq c < 0$. Since this holds for any lift, $c \notin \mathbb{Z}$ (since if it were in $\mathbb{Z}$, we could choose the lift which makes it exactly 0, contradicting our arguments above). Thus, $\rho(g) \neq 0$. □

LEMMA 6.1.4. *Let $g : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ be a degree one homeomorhpism, and $\tilde{g}$ be a lift of g. Let $G = \tilde{g}^q - p$. Then $\rho_0(G) = q\rho_0(\tilde{g}) - p$.*

PROOF. We saw in the proof of consistency of the definition of ρ that $\rho_0(\tilde{g} + k) = \rho_0(\tilde{g}) + k$ (this is why for rotation numbers on the circle, we take everything mod 1). Now observe that:

$$\rho_0(\tilde{g}^q) = \lim_{n\to\infty}\frac{\tilde{g}^{nq}(x) - x}{n} = q\lim_{n\to\infty}\frac{\tilde{g}^{nq}(x) - x}{nq} = q\rho_0(\tilde{g})$$

□

PROOF OF PROPOSITION 6.1.2. Let g be a homeomorphism and $\tilde{g}$ be a lift with $\rho_0(\tilde{g}) = \frac{p}{q}$. Then $\rho_0(\tilde{g}^q - p) = 0$, so $\tilde{g}^q - p$ has a fixed point by Lemma 6.1.3. That is, there is some $x \in \mathbb{R}$ such that $\tilde{g}^q(x) = x + q$.

Now suppose that $\tilde{g}^k(y) = y + l$ for some $y \in \mathbb{R}$ and $k, l \in \mathbb{Z}$ (ie, $[y]$ is periodic for g with period k). Then $\tilde{g}^k - l$ has a fixed point, so $\rho_0(\tilde{g}^k - l) = 0$ and $\rho_0(\tilde{g}) = k/l$. So $k/l = p/q$ and since p and q are coprime, there is an integer m such that $k = mp$ and $l = mp$. So q is in fact the smallest period for the map g, and every other period is a multiple of q. □

We state the following Proposition for reference, it will appear in the homework:

PROPOSITION 6.1.5. *If $f, g, h : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ are homeomorphisms of degree 1 and $f \circ h = h \circ g$, then $\rho(f) = \rho(g)$*

Again, let $\mathcal{H}(\mathbb{R}/\mathbb{Z})$ be the sets of homeomorphisms of $\mathbb{R}/\mathbb{Z}$, and recall that we can put a metric (this is often called the *uniform metric* or the sup-*metric*) on $\mathcal{H}(\mathbb{R}/\mathbb{Z})$ by setting:

$$d(f, g) = \sup_{x \in \mathbb{R}/\mathbb{Z}} d(f(x), g(x)) = \max_{x \in \mathbb{R}/\mathbb{Z}} d(f(x), g(x))$$

We need one final lemma before stating the proposition:

LEMMA 6.1.6. *Let X be a compact metric space, and $k \in \mathbb{N}$. The map $e_k : \mathcal{H}(X) \to \mathcal{H}(X)$ defined by $e_k(f) = f^k$ is a continuous map with respect to the uniform metric*

PROOF. We will show the ε-δ definition of continuity. Let $\varepsilon > 0$, $f \in \mathcal{H}(X)$. We proceed by induction on k. The base case of $e_1 = \mathrm{id}$ is obvious. Now we proceed to the inductive step. Let $\delta_0 > 0$ be such that if $d(x, y) < \delta_0$, then $d(f(x), f(y)) < \varepsilon/2$. Let $\delta > 0$ be such that if $d(f, g) < \delta$, then $d(f^{k-1}(x), g^{k-1}(x)) < \min\{\varepsilon/2, \delta_0\}$. Then for every $x \in X$:

$$d(g^k(x), f^k(x)) \leq d(g(g^{k-1}(x)), f(g^{k-1}(x))) + d(f(g^{k-1}(x)), f(f^{k-1}(x))) < \varepsilon/2 + \varepsilon/2 = \varepsilon$$

Then $d(g^k, f^k) < \varepsilon$. □

PROPOSITION 6.1.7. *The map $\rho : \mathcal{H}(\mathbb{R}/\mathbb{Z}) \to \mathbb{R}/\mathbb{Z}$ is continuous with respect to the uniform metric*

PROOF. In the proof of Lemma 6.1.3, we saw that if $u(x) = \tilde{g}(x) - x \geq c$, then $\rho_0(\tilde{g}) \geq c$ (and similarly, if $u(x) \leq c$, then $\rho_0(\tilde{g}) \leq c$). In fact, the converse is true for a power: if $\rho_0(\tilde{g}^q) > p$ (that is, $\rho_0(\tilde{g}) > p/q$), then there exists a $\delta > 0$ such that $u(x) \geq p + \delta$, where $u(x) = \tilde{g}^q(x) - x$.

To prove this, consider again $G = \tilde{g}^q - p$. Then $\rho_0(G) = q\rho_0(\tilde{g}) - p > 0$. Then G has no fixed points, and $U(x) = G(x) - x$ is positive everywhere, so there exists a $\delta > 0$ such that $U(x) \geq \delta$. That is, $u(x) = \tilde{g}^q(x) - x \geq p + \delta$.

We will show that the inverse image of rational intervals is open in $\mathcal{H}(\mathbb{R}/\mathbb{Z})$. Suppose for some lift $\tilde{g}$ of g,$\rho_0(\tilde{g}) \in (p_1/q_1, p_2/q_2)$. Then by the above arguments there exists a $\delta > 0$ such that if $u_1(x) = \tilde{g}^{q_1}(x) - p_1$ and $u_2(x) = \tilde{g}^{q_2}(x) - p_2$, $u_1(x) \geq \delta$ and $u_2(x) \leq \delta$. Choose ε such that if $d(f, g) < \varepsilon$, $d(f^k, g^k) < \delta$ for $k = q_1, q_2$. In this case, if $w_i(x) = \tilde{f}^{q_i}(x) - p_i$, then $w_1 > 0$ and $w_2 < 0$. In particular, $\rho_0(\tilde{f}) > p_1/q_1$ and $\rho_0(\tilde{f}) < p_2/q_2$. That is, $\rho_0(B_\varepsilon(\tilde{g})) \subset (p_1/q_1, p_2/q_2)$, and ρ_0 (and hence ρ) is continuous. □

COROLLARY 6.1.8. *Let $f : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ be a degree one homeomorhpism. For $t \in \mathbb{R}/\mathbb{Z}$, define $f_t(x) = f(x) + t$. Suppose that $\rho(f_{t_0}) \neq \rho(f_{t_1})$. Then there exists some $t \in (t_0, t_1)$ such that f_t has a periodic point (ie, $\rho(f_t)$ is rational) and some $s \in (t_0, t_1)$ such that f_s has no periodic points (ie, $\rho(f_s)$ is irrational). Moreover, the map $t \mapsto \rho_0(\tilde{f}_t)$ is increasing for a continuous family of lifts of f_t*

PROOF. That $t \mapsto \rho_0(\tilde{f}_t)$ is increasing follows from the fact that if $s < t$, then $\tilde{f}_s(x) < \tilde{f}_t(x)$ and if $u_t(x) = \tilde{f}_t(x) - x$, $u_s(x) < u_t(x)$ so:

$$\rho_0(\tilde{f}_t) = \lim_{n\to\infty} \frac{1}{n}\left(\tilde{f}_t^n(x) - x\right) = \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} u_t(x) \geq \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} u_s(x) = \rho_0(\tilde{f}_s)$$

The rest follows from the fact that $t \mapsto f_t$ is continuous with respect to the uniform metric, so continuity of ρ_0 and the intermediate value theorem implies the result. □

6.2. Irrational Rotation Number

Let $f : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ be a homeomorphism of degree 1 with $\rho(f) \notin \mathbb{Q}$, and $\tilde{f} : \mathbb{R} \to \mathbb{R}$ be a lift (so $\rho_0(\tilde{f}) \notin \mathbb{Q}$). Recall that $\rho_0(\tilde{f}) < p/q$ if and only if $\tilde{f}^q(x) - x - p < 0$ everwhere, and if at some x_0, $\tilde{f}^q(x_0) - x_0 - p < 0$, then $q\rho_0(\tilde{f}) - p \leq 0$. Since we have assumed that $\rho_0(\tilde{f}) \notin \mathbb{Q}$, we can conclude that $\rho_0(\tilde{f}) < p/q$ if and only if $\tilde{f}^q(x) - x - p < 0$ for every (or for some) $x \in \mathbb{R}$. We get the following Corollary:

Corollary 6.2.1. *Suppose that $\rho(f) \notin \mathbb{Q}$. If $n_1, n_2, m_1, m_2 \in \mathbb{Z}$, and $x \in \mathbb{R}$:*

$$\tilde{f}^{n_1}(x) + m_1 < \tilde{f}^{n_2}(x) + m_2 \text{ if and only if } n_1\rho_0(\tilde{f}) + m_1 < n_2\rho_0(\tilde{f}) + m_2$$

Proof. Assume that $n_1 \geq n_2$. Then if $y = \tilde{f}^{n_2}(x)$:

$$\tilde{f}^{n_1-n_2}) - y - (m_2 - m_1) < 0 \text{ if and only if } (n_1 - n_2)\rho_0(\tilde{f}) - (m_2 - m_1) < 0$$

□

The best outcome of these systems would be a topological conjugacy h satisfying $h \circ f = R_{\rho(f)} \circ h$. Unfortunately, this is not always true, but with a little work, we can get a complete picture of what happens. Let $A = \left\{\tilde{f}^q(0) + p : q, p \in \mathbb{Z}\right\}$, so that $\mathcal{O}_f(0) = p(A)$. Furthermore, let $B = \left\{q\rho_0(\tilde{f}) + p : p, q \in \mathbb{Z}\right\}$, so that $\mathcal{O}_{R_{\rho(f)}} = p(B)$. B is dense in $\mathbb{R}$ since rotation by an irrational number has a dense orbit. Define the function $H : A \to B$ by setting $H(\tilde{f}^q(0) + p) = q\rho_0(\tilde{f}) + p$. Note that this is well-defined since if $x = \tilde{f}^{q_1}(0) + p_1 = \tilde{f}^{q_2}(0) + p_2$. In this case, if $q_1 \neq q_2$, x would be a periodic point for f (and we have assumed that f has irrational rotation number), so we conclude that $q_1 = q_2$ and hence $p_1 = p_2$.

A is an increasing map, since if $z = F^{n_1}(0) + m_1 < F^{n_2}(0) + m_2 = w$, so Corollary 6.2.1 gives us that H is increasing. Now we have the following properties:

(i) B is dense in $\mathbb{R}$
(ii) $\tilde{f}(A) = A$ and $R_{\rho(f)}(B) = B$
(iii) H is increasing
(iv) $H(x + 1) = H(x) + 1$
(v) $H \circ \tilde{f} = R_{\rho_0(\tilde{f})} \circ H$

So we need only to make the following observation.

Proposition 6.2.2. *If $A, B \subset \mathbb{R}$, $\overline{B} = \mathbb{R}$ and $H : A \to B$ is increasing and onto, then there is a unique continuous map $h : \mathbb{R} \to \mathbb{R}$ such that $h|_A = H$. Moreover, H is increasing. In particular, for $z \in \mathbb{R}$, $h^{-1}(z)$ is either a single point or a closed interval.*

Proof. Our candidate function is:

$$h(x) = \sup_{\substack{y \leq x \\ y \in A}} H(y)$$

Then h is clearly increasing. To see that it is continuous, suppose that h was discontinuous at a point x. Then since H is increasing the only discontinuity that can appear is a jump discontinuity:

$$L = \lim_{t \to x^+} h(t) = \inf_{t > x} h(t) = \inf_{t > x} \sup_{\substack{s \leq t \\ s \in A}} H(s) = \inf_{\substack{t > x \\ t \in A}} H(t) > H(x)$$

But then $H(\mathbb{R}) \not\supset (H(x), L)$, so H cannot be onto a dense set. □

These types of functions can have very strange. One such example is a *Devil's Staircase*. We can use Proposition 6.2.2 to find its description. Indeed, let $A = \{\sum_{k=1}^n \alpha_n 3^{-n} : \alpha_n = 0, 2\}$, and $B = \{\sum_{k=1}^n \beta_b 2^{-n}\}$. Then define $H : A \to B$ by sending $x = \sum \alpha_n 3^{-n} \mapsto \sum \alpha_n 2^{-n}$. This map is obviously increasing and onto and B is dense. The graph of the function on $[0, 1]$ is plotted in Figure 3.

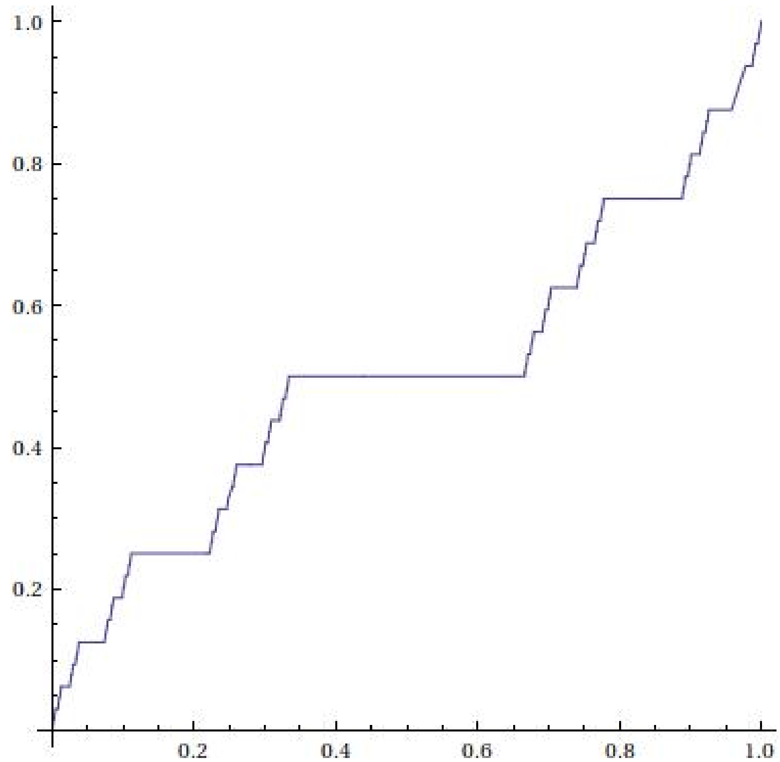

FIGURE 3. The Standard Devil's Staircase

Let us make one final remark here. The devil's staircase is also a standard counterexample in analysis: it is an increasing function, and hence is differentiable on a set of full measure (also, in a topological analogy, on a dense-G_δ set). But one quickly sees that this derivative is 0 (it is constant on the removed intervals of the cantor set)! But then:

$$1 = h(1) - h(0) \stackrel{?}{=} h(0) + \int_0^1 h'(t)\, dt = h(0)$$

Which is quickly incorrect. What fails here is *absolute continuity*, a notion which you can find in most real analysis textbooks.

COROLLARY 6.2.3. *If $\rho(f) \notin \mathbb{Q}$, there exists a continuous, increasing, degree 1 semiconjugacy h such that $h \circ f = R_{\rho(f)} \circ h$. In particular, $h^{-1}(z)$ is either a point or a closed interval.*

To finish our discussion of irrational rotation nuber, let $S = \left\{z \in \mathbb{R}/\mathbb{Z} : \left|h^{-1}(z)\right| > 1\right\}$. Then let $U = \bigcup_{z \in S} \operatorname{Int}(h^{-1}(z))$, so that U is a union of open intervals. Then U is invariant by f, since h is a semiconjugacy, $\Lambda_f = \mathbb{R}/\mathbb{Z} \setminus U$ is a compact invariant subset, and:

PROPOSITION 6.2.4. *If $x \in \mathbb{R}/\mathbb{Z}$, $\omega_f(x) = \alpha_f(x) = \Lambda_f$*

PROOF. Let $x \in \mathbb{R}/\mathbb{Z}$. First, if $x \in U$, then $x \in h^{-1}(z_0)$ for some z_0. Since there are no periodic points and U is invariant, we get that $f^n(x) \in h^{-1}(z_n)$ for pairwise distinct z_n. But since $\operatorname{diam} h^{-1}(z_n) \to 0$ (otherwise, our circle would have infinite length), we get that $f^n(x)$ follows the orbit of the boundaries of $h^{-1}(z_n)$ closely. Hence $\omega_f(x) = \omega_f(z')$, where z' is either of the boundary points of $\mathbb{R}/\mathbb{Z}$.

So we need only to investigate the orbits of f in Λ_f. But if $x \in \Lambda_f$, $h(\omega_f(x)) = \omega_{R_{\rho(f)}}(h(x)) = \mathbb{R}/\mathbb{Z}$. Hence, for every $z \in \mathbb{R}/\mathbb{Z}$, $h^{-1}(z) \cap \omega_f(x) \neq \emptyset$. Since $h^{-1}(z)$ is either an interval or a point, and $\omega_f(x)$ is closed, $\omega_f(x) \supset \mathbb{R}/\mathbb{Z} \setminus U = \Lambda_f$. The other inclusion is obvious. $\square$

Corollary 6.2.5. *If $f : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ has $\rho(f) \notin \mathbb{Q}$, $f|_{\Lambda_f}$ is minimal*

We state the following proposition and leave the proof as an exercise:

Proposition 6.2.6. *Let f be such that $\rho(f) \notin \mathbb{Q}$. The following are equivalent:*

(i) *f is topologically transitive*
(ii) *f is minimal*
(iii) $\Lambda_f = \mathbb{R}/\mathbb{Z}$
(iv) *f is topologically conjugate to $R_{\rho(f)}$*

Exercise 6.2.1. Show that Λ_f is either all of $\mathbb{R}/\mathbb{Z}$ or contains no intervals. [*Hint*: Show that if Λ_f contains an interval, the interval's boundary points are periodic]

Note that for every rigid rotation, $\Lambda_f = \mathbb{R}/\mathbb{Z}$. One can construct maps which have $\Lambda_f \neq \mathbb{R}/\mathbb{Z}$ using the following process. Let us first describe it intuitively. As we are describing maps on the circle, let us imagine our circle is an infinitely thin band of wire. We pick intervals $[a_n, b_n]$ along the wire for $n \in \mathbb{Z}$, and at each point, pinch the wire so that the length of $[a_n, b_n]$ forms a seperate loop (See Figure 4). Now, if we make the length of the intervals decay very quicly (say $b_n - a_n \leq 2^{-(n+1)}$), we can remove them and be left again with a circle. On this remaining circle, we act by an irrational rotation (or some conjugate of it, so that the "pinch points" form a single irrational orbit, see Figure 5). Then, simply map each of the pinched intervals linearly to the next in the image. It is clear that no point is periodic, so the rotation number is irrational. Furthermore, the interval length decays to 0, so any point in an interval will tend to a point on the boundary. Note that in this case, $\Lambda_f = \mathbb{R}/\mathbb{Z} \setminus \bigcup_n (a_n, b_n)$.

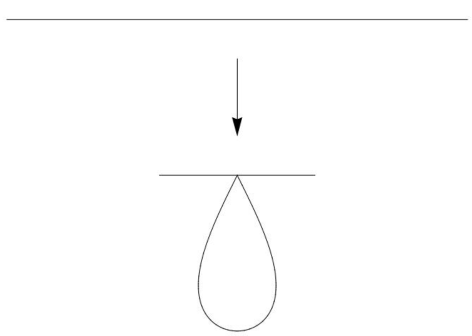

Figure 4. Pinching a single interval

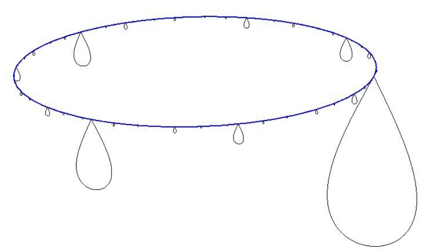

Figure 5. The circle with many pinched intervals

6.3. The Denjoy Theorem

If we insist our homeomorphism is nice (as in, it is sufficiently differentiable), these problems disappear:

Definition 6.3.1. A map $f : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ is a *diffeomorphism* if f has a lift $\tilde{f}$ such that $\tilde{f}$ is invertible, and $\tilde{f}$, $\tilde{f}^{-1}$ are differentiable

We will often use the notation that $f'([x]) = \tilde{f}'(x)$. Since two lifts of f will differ by a constant, the derivative of f is independent of the choice of lift. This also allows us to speak of higher-order derivatives, using the same language.

THEOREM 6.3.2 (Denjoy). *If $f : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ is a C^2 diffeomorphism and $\rho(f) \notin \mathbb{Q}$, $\Lambda_f = \mathbb{R}/\mathbb{Z}$. That is, the h constructed in Corollary 6.2.3 is a homeomorphim.*

We remark that it suffices to take $f \in C^1$ such f' has *bounded variation*. C^2 maps are the simplest example of this.

PROOF. Suppose that $\Lambda_f \neq \mathbb{R}/\mathbb{Z}$, and choose $x \in \mathbb{R}/\mathbb{Z}$ such that $h^{-1}(x) = I = [a, b]$ with $a \neq b$. We wish to arrive at a contradiction. We first note that $f^k(I) \cap f^l(I) = \emptyset$ if $k \neq l$. This follows because $f^n(I) = f^n(h^{-1}(x)) = h^{-1}(R^n_{\rho(f)}(x))$ and $R_{\rho(f)}$ has no periodic points, so:

$$f^k(I) \cap f^l(I) = h^{-1}(R^k_{\rho(f)}(x)) \cap h^{-1}(R^l_{\rho(f)}(x)) = h^{-1}\left(\left\{R^k_{\rho(f)}(x)\right\} \cap \left\{R^l_{\rho(f)}(x)\right\}\right) = \emptyset$$

Now, if $J = [a, b] \subset \mathbb{R}/\mathbb{Z}$ is an interval, set $|J| = b - a$. This implies that since the iterates of I are disjoint:

$$\sum_{n\in\mathbb{Z}} |f^n(I)| \leq 1$$

In particular, $\sum_{n\geq 0} |f^n(I)| < 1$. We use the fact that this is convergent to get that for any δ_0 there exists an N such that $n \geq N$, $\sum_{n\geq N} |f^n(I)| < \delta_0$. By choosing $x' = R^N_{\rho(f)}(x)$ rather than x', we can make $\sum_{n\geq 0} |f^n(I)|$ arbitrarily small by the appropriate choice of x'.

We will use this condition to prove the following Lemma.

LEMMA 6.3.3. *Let $f : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ be a C^2 diffeomorphism, and I a nontrivial interval. Then there are $\delta_0 > 0$ and $\varepsilon > 0$ such that if*

$$\sum_{n\geq 0} |f^n(I)| \leq \delta_0$$

then

$$\sum_{n\geq 0} |f^n(I_\varepsilon)| \leq 2 \sum_{n\geq 0} |f^n(I)|$$

Where I_ε is an interval containing I, and $|I_\varepsilon| / |I| = 1 + \varepsilon$, e.g. if $I = [a, b]$, $I_\varepsilon = [a - r, b + r]$, where $r = \varepsilon(b - a)/2$.

Indeed we can take $\varepsilon = 1/2$ in the Lemma.

Let us first use the Lemma to finish the proof. It immediately implies that $|f^n(I_\varepsilon)| \to 0$. We know that if $I = [a, b]$, $a \in \Lambda_f$, so $a \in \omega(a) = \Lambda_f$. Now expand the interval I to I_ε so that ε is very small, and $I_\varepsilon = [a - r, b + r]$. Now we choose n_i sufficiently large so that $|f^{n_i}(I_\varepsilon)| < r/2$. Furthermore, we can assume that $d(f^n(a), a) < r/2$ by choosing n_i sufficiently large.

But if $y \in f^{n_i}(I_\varepsilon)$, $d(y, f^{n_i}(a)) < r/2$, and $d(y, a) \leq d(y, f^{n_i}(a)) + d(f^{n_i}(a), a) < r$. So $\overline{f^{n_i}(I_\varepsilon)} \subset I_\varepsilon$. But then f^{n_i} has a fixed point in I_ε and $\rho(f) \in \mathbb{Q}$. □

One may note that nowhere in the proof did we use that $f \in C^2$. This will appear in the proof of the Lemma:

PROOF OF LEMMA. We will prove that for some $\varepsilon > 0$ and $\delta_0 > 0$,

$$\sum_{n=0}^{N} |f^n(I_\varepsilon)| \leq 2 \sum_{n=0}^{N} |f^n(I)|$$

for every $N \geq 0$. This proves the Lemma since we can then send N to ∞. We will induct over N. In the case when $N = 0$, this amounts to choosing $\varepsilon > 0$ such that $|I_\varepsilon| \leq 2 |I|$, which is obviously possible by taking $\varepsilon < 1$ (so $\varepsilon = 3/2$ works).

Now we prove the inductive step. Note that:

$$\sum_{n=0}^{N} |f^n(I_\varepsilon)| = \sum_{n=0}^{N-1} |f^n(I_\varepsilon)| + \left|f^N(I_\varepsilon)\right| \leq 2\sum_{n=0}^{N-1} |f^n(I)| + \left|f^N(I_\varepsilon)\right|$$

So we need only to show that $\left|f^N(I_\varepsilon)\right| \leq 2\left|f^N(I)\right|$. Here we see derivatives appear, as this inequality is the same as:

$$f'(y_N) = \frac{\left|f^N(I_\varepsilon)\right|}{|I_\varepsilon|} \leq \frac{2}{1+\varepsilon}\frac{\left|f^N(I)\right|}{|I|} = f'(z_N)$$

for some points $y_N \in I_\varepsilon$, $z_N \in I$, by the mean value theorem. Since we have assumed that f is a degree 1 diffeomorphism, $f'(x) > 0$ everywhere. Furthermore,

$$(f^n)'(x) = f'(f^{n-1}(x)) \cdot (f^{n-1})'(x) = \prod_{k=0}^{n-1} f'(f^k(x)).$$

Taking logarithms, $\log(f^n)'(x) = \sum_{k=0}^{n-1} \log f'(f^k(x))$. Now, since $f'(x) > 0$, $\log f' \in C^1$. Thus, there exists $L \geq 0$ such that $|\log f'(x) - \log f'(y)| \leq Ld(x,y)$ (L is called the Lipschitz constant, and in our case, is $\sup \frac{f''(x)}{f'(x)}$). Let us compute:

$$\begin{aligned}
\left|\log\frac{(f^N)'(x)}{(f^N)'(y)}\right| &= \left|\sum_{k=0}^{N-1} \log f'(f^k(x)) - \log f'(f^k(y))\right| \leq \sum_{k=0}^{N-1} \left|\log f'(f^k(x)) - \log f'(f^k(y))\right| \\
&\leq \sum_{k=0}^{N-1} Ld(f^k(x), f^k(y))
\end{aligned}$$

Note that L is chosen independently of N. Notice also that if $x, y \in I_\varepsilon$, then $f^k(x), f^k(y) \in f^k(I_\varepsilon)$ and hence $d(f^k(x), f^k(y)) \leq \left|f^k(I_\varepsilon)\right|$ for every k. Hence we get that

$$\left|\log\left((f^N)'(x)/(f^N)'(y)\right)\right| \leq L\sum_{k=0}^{N-1}\left|f^k(I_\varepsilon)\right| \leq 2L\sum_{k=0}^{N-1}\left|f^k(I)\right| \leq 2L\delta_0$$

where the second inequality follows from induction and the third by hypothesis (we will fix δ_0 soon).

Now, removing the logarithms, we get:

$$(f^N)'(x) \leq (f^N)'(y)\exp(2L\delta_0)$$

for every pair of points $x, y \in I_\varepsilon$. Chose $x_N \in I_\varepsilon$ and $y_N \in I \subset I_\varepsilon$ such that

$$\frac{\left|f^N(I_\varepsilon)\right|}{|I_\varepsilon|} = (f^N)'(x_N) \qquad \text{and} \qquad \frac{\left|f^N(I)\right|}{|I|} = (f^N)'(y_N).$$

Then we get that

$$\begin{aligned}
\left|f^N(I_\varepsilon)\right| &= \frac{\left|f^N(I_\varepsilon)\right|}{|I_\varepsilon|}(1+\varepsilon)\,|I| \\
&= (f^N)'(x_N)(1+\varepsilon)\,|I| \\
&\leq (f^N)'(y_N)\exp(2L\delta_0)(1+\varepsilon)\,|I| \\
&= \frac{\left|f^N(I)\right|}{|I|}\exp(2L\delta_0)(1+\varepsilon)\,|I| \\
&= \exp(2L\delta_0)(1+\varepsilon)\left|f^N(I)\right|
\end{aligned}$$

Now, we simply choose $\varepsilon = 1/2$ and δ_0 small enough so that

$$\frac{3}{2}\exp(2L\delta_0) \leq 2$$

(note that this choice is independent of N). Then we get exactly the desired inequality.

□

6.4. Rational Rotation Number

Let us now discuss the case when $\rho(f) = p/q \in \mathbb{Q}$, written in reduced form (ie, $(p, q) = 1$). By Proposition 6.1.4, we know that for an appropriate lift $\tilde{f}$ of f, there is some point $x \in \mathbb{R}$ such that $\tilde{f}^q(x) = x + p$. Furthermore, if $\tilde{f}^k(y) = y + l$, then $l/k = p/q$. We claim that in this case, $k = q$ and $l = p$ (we want to rule out, for instance, $k = 2q$ and $l = 2p$). This will follow from the following definition:

Proposition 6.4.1. *Let $g : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ be a degree one homeomorphism and $\tilde{g} : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ have $\rho_0(\tilde{g}) = 0$. Then there are two orbit types:*

(i) x is fixed
(ii) $\alpha(x) = \{y_1\}$, $\omega(x) = \{y_2\}$, where y_1 and y_2 are fixed points

Corollary 6.4.2. *If $\rho_0(\tilde{g}) = p/q$, then periodic points for g lift to points of $\mathbb{R}$ such that $\tilde{g}^q(x) - p = x$. Furthermore, every periodic point has the same period, and if a point $x \in \mathbb{R}/\mathbb{Z}$ is not periodic, $\omega(x) = \mathcal{O}(y_+)$ and $\alpha(x) = \mathcal{O}(y_-)$, where $y_\pm$ is some periodic point.*

We leave it as an exercise to the reader to deduce the corollary from the proposition. To prove Proposition 6.4.1, we will prove the following Lemma:

Lemma 6.4.3. *If $f : [0, 1] \to [0, 1]$ is a strictly increasing homeomorphism, then orbit types of f fall into the same dichotomy of Proposition 6.4.1*

Proof. Let F denote the set of fixed points of f, and note that F is a closed subset of $[0, 1]$, since $F = (f - \mathrm{id})^{-1}(0)$. Then $W = [0, 1] \setminus F$ is open and hence a union of disjoint open intervals $W = \bigcup_\alpha U_\alpha = \bigcup_\alpha (a_\alpha, b_\alpha)$. Furthermore, a_α and b_α are fixed for every α. Finally, on each such interval, $f - \mathrm{id}$ is either strictly greater than or less than 0, and $f(U_\alpha) = U_\alpha$ (this is because the endpoints are fixed and f is strictly increasing).

Now, if $y \in U_\alpha$, and $f(y) > y$, then since the orbit must lie in the interval, and on this interval, $f - \mathrm{id} > 0$. Thus:

$$\dots f^n(y) > f^{n-1}(y) > \dots > f^2(y) > f(y) > y > f^{-1}(y) > f^{-2}(y) > \dots$$

Since $b_n = f^n(y)$ is an increasing sequence, and bounded, it must converge. It converges to a fixed point, since if β is the limit:

$$f(\beta) = f\left(\lim_{n\to\infty} f^n(y)\right) = \lim_{n\to\infty} f^{n+1}(y) = \beta$$

and that fixed point must be b_α (since it lies in $\overline{U_\alpha} = [a_\alpha, b_\alpha]$ and $y > \beta$). Similarly, if $a_n = f^{-n}(y)$, this is a decreasing sequence, and hence must also converge. It again converges to a fixed point, so by a parallel argument, it must be a_α. □

Now we are ready to prove Proposition 6.4.1:

Proof. Let $\tilde{g} : \mathbb{R} \to \mathbb{R}$ be a lift with a fixed point, call it x_0. Then if we restrict $\tilde{g}|_{[x_0,x_0+1]}$ to get a strictly increasing homeomorphism. We can apply Proposition 6.4.3 to get that the orbit types are exactly as perscribed. Now, we note that on each interval $[x_0 + k, x_0 + k + 1]$, the behavior is the same. Thus since $\mathbb{R} = \bigcup_{k\in\mathbb{Z}} [x_0 + k, x_0 + k + 1]$, $\tilde{g}$ has these orbits everywhere. □

We have one final observation describing the rational rotation number setting. Recall that if $[x] \in \mathbb{R}/\mathbb{Z}$:

$$\tilde{f}^{m_1}(x) - x - n_1 \le \tilde{f}^{n_2}(x) - x - m_2 \text{ if and only if } m_1\rho_0(\tilde{f}) - n_1 \le m_2\rho_0(\tilde{f}) - n_2$$

(We proved a stronger version of this using strict inequalities for irrational rotation number) This is often interpreted as "preserving the combinatorics." That is, the intervals between periodic points are permuted in the same way that they are for the rational rotation, given the appropriate number of periodic orbits.

CHAPTER 7

Moving to Higher Dimensions: Three Prototypes

So far we have focused mainly on dynamics on the circle, but of course, there are many other spaces which support a dynamical system. We present a few examples, which like the maps L_d and R_α, will motivate our studies. They will be presented in 3 seperate sections:

7.1. Horseshoes

Consider $I^2 = [0,1]^2 \subset \mathbb{R}^2$, and a map $g : I^2 \to \mathbb{R}^2$ which has the following picture:

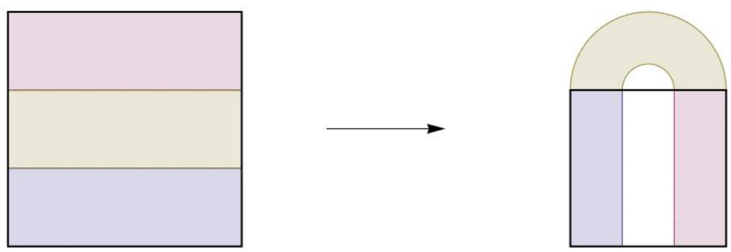

Figure 1. Playing Horseshoes

That is, f linearly maps the blue (bottom) region to the blue region on the right figure, preserving the bottom right corner, and maps the red (top) region to the red region on the right figure, mapping the top left corner to the bottom right (see Equation (2)). The yellow (center) region is mapped to the "band" at the top so that the map f is continuous.

Like in the case of the map $L_3 : [0,1/3] \cup [2/3,1] \to [0,1]$, the forward (and in this case, backward) orbits are also not well-defined for certain points which do not lie in $I^2 \cap g^{-1}(I^2)$ (or in the case of backward orbits, $I^2 \cap g(I^2)$).

The preimage of the "band" at the top of the horseshoe (that is, $g(I^2) \setminus I^2$) will be a horizontal strip in the center of the square. Removing this preimage, we get two disjoint horizontal rectangles, R_0 and R_1. Let us assume that $R_0 = [0,1] \times [0,1/3]$ and $R_1 = [0,1] \times [2/3,1]$. Let us also set $f : R_0 \times R_1 \to [0,1] \times [0,1]$ as:

$$f(x,y) = \begin{cases} (x/3, 3y) & (x,y) \in R_0 \\ (1 - x/3, 3 - 3y) & (x,y) \in R_1 \end{cases} \tag{2}$$

f is then the restrictiion of a Horseshoe, but we still call f a Horseshoe as it shares the fundamental properties. Now, we set $\Lambda_f = \bigcap_{n\in\mathbb{Z}} f^n(R_0 \cup R_1) = \bigcap_{n\in\mathbb{Z}} f^n(I^2)$ (here we need to take negative values of n as well because f is not surjective). Then $f : \Lambda_f \to \Lambda_f$ is an *invertible* dynamical system. We could equivalently define:

$$\Lambda_f^+ = \bigcap_{n\leq 0} f^n(I^2) \tag{3}$$

$$\Lambda_f^- = \bigcap_{n\geq 0} f^n(I^2) \tag{4}$$

Then $f : \Lambda_f^+ \to \Lambda_f^+$ is a noninvertible dynamical system, as is $f^{-1} : \Lambda_f^- \to \Lambda_f^-$. Furthermore, $\Lambda_f = \Lambda_f^+ \cap \Lambda_f^-$.

We can find explicit descriptions for Λ_f, Λ_f^+ and Λ_f^-, due to the simple description of the dynamics f. Let us first focus on Λ_f^+. Note that both strips R_0 and R_1 are mapped linearly to vertical strips $[0, 1/3] \times [0, 1]$ and $[2/3, 1] \times [0, 1]$. In this case, the points whose images do *not* lie in $R_0 \cup R_1$ are exactly those points which lie in the center strip of R_0 and R_1. Continuing this process, we remove thinner and thinner strips, ending up with the product of a Cantor set with $[0, 1]$. That is, if C is the standard Cantor set, then $\Lambda_f^+ = [0, 1] \times C$. See Figure 2.

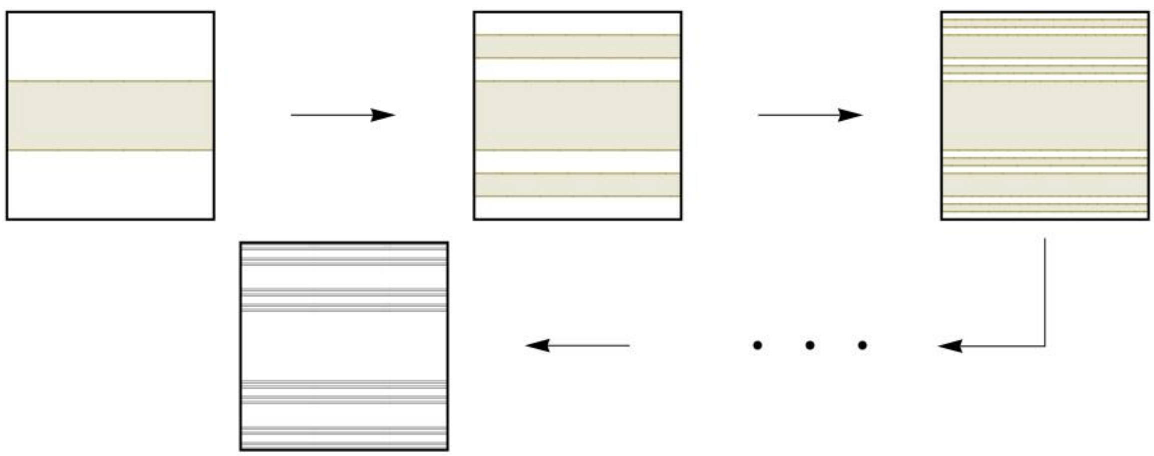

FIGURE 2. Construction of Λ_f^+

Similarly, $\Lambda_f^- = C \times [0, 1]$, so $\Lambda_f = C \times C$.

This example also serves to establish prototypical structures that we see appear many times in higher dimensions. Define for $x \in \Lambda_f$:

$$\begin{aligned}
W^s(x) &= \Big\{y \in \Lambda_f^+ : d(f^n(x), f^n(y)) \to 0 \text{ as } n \to \infty\Big\} \\
W^s_\varepsilon(x) &= \Big\{y \in \Lambda_f^+ : d(f^n(x), f^n(y)) \leq \varepsilon \text{ for all } n \geq 0\Big\} \\
W^u(x) &= \Big\{y \in \Lambda_f^- : d(f^n(x), f^n(y) \to 0 \text{ as } n \to -\infty\Big\} \\
W^u_\varepsilon(x) &= \Big\{y \in \Lambda_f^- : d(f^n(x), f^n(y)) \leq \varepsilon \text{ for all } n \leq 0\Big\}
\end{aligned}$$

We state the following Proposition without proof. The proof may appear later, but the notion of local and global *stable and unstable manifolds* are crucial to the study of the behavior of dynamical systems in higher dimensions:

PROPOSITION 7.1.1. *For the map $f : R_0 \cup R_1 \to I^2$ and $x \in \Lambda_f$:*

(i) *For sufficiently small $\varepsilon > 0$: $W^s_\varepsilon(x) \subset W^s(x)$*

(ii) *For sufficiently small* $\varepsilon > 0$*:* $W^u_\varepsilon(x) \subset W^u(x)$

(iii) *For sufficiently small* $\varepsilon > 0$*:*

$$W^s(x) = \bigcup_{n\geq 0} f^{-n}(W^s_\varepsilon(f^n(x)) = \left\{y \in \Lambda_f^+ : f^n(y) \in W^s_\varepsilon(f^n(x)) \text{ for some } n \geq 0\right\}$$

(iv) *For sufficiently small* $\varepsilon > 0$*:*

$$W^u(x) = \bigcup_{n\geq 0} f^{n}(W^u_\varepsilon(f^{-n}(x)) = \left\{y \in \Lambda_f^- : f^n(y) \in W^u_\varepsilon(f^n(x)) \text{ for some } n \leq 0\right\}$$

(v) $\overline{W^s(x)} = \Lambda_f^+$

(vi) $\overline{W^u(x)} = \Lambda_f^-$

To find a symbolic description of Horseshoes, we again use shift spaces. Because $\sigma : \Sigma_d^+ \to \Sigma_d+$ is *not* invertible, we are forced to consider another space:

Definition 7.1.2. For $d \geq 2$, the *two-sided shift space* on d symbols is the space $\Sigma_d = \{1, \dots, d\}^{\mathbb{Z}} = \{\mathbf{x} : \mathbf{x} = (\dots, x_{-2}, x_{-1}, x_0, x_1, \dots)\}$. Then define $\sigma : \Sigma_d \to \Sigma_d$ by setting $\sigma(\mathbf{x})_n = x_{n+1}$. Put a topology on Σ_d by setting;

$$d_\beta(\mathbf{x}, \mathbf{y}) = \sum_{k=-\infty}^{\infty} |x_k - y_k| \, \beta^{|k|}$$

For some $0 < \beta < 1$

We leave it as an exercise to verify that Σ_d is again compact, and that the topology is independent of β. Note that in this case, however σ is invertible, with $\sigma^{-1}(\mathbf{x})_n = x_{n-1}$. Our plan will be similar to that of Markov partitions (see Section 5.2). That is, we wish to construct a map $h : \Sigma_d \to \Lambda_f$ such that $h(\mathbf{x}) = y$ if and only if $f^i(y) \in R_{x_i}$ for every $i \in \mathbb{Z}$. Note that because $f : \Lambda_f \to \Lambda_f$ is invertible, this allows us to consider $i \in \mathbb{Z}$ rather than just $i \in \mathbb{N}_0$. We wish to define:

$$h(x) = \bigcap_{i\in\mathbb{Z}} f^i(R_{x_i})$$

Proposition 7.1.3. *For the map* $f : \Lambda_f \to \Lambda_f$*,* h *is well defined and a homeomorphism, conjugating* f *to the two-sided shift.*

Sketch of Proof. We only outline the steps of the proof here. We observed before that Λ_f^+ is exactly $I \times C$. Furthermore, if we only look at the second component of the map, at each step, we either applied $y \mapsto 3y$ or $y \mapsto 3 - 3y$. In the same way as in the markov partition, we see that each infinite forward sequence gives a unique horizontal strip.

Now we repeat the argument with the inverse, defined on the rectangles $f(R_0)$ and $f(R_1)$. This time, we get for each infinite backward sequence, a unique vertical strip. Now simply define h of the bidirectional sequence to be the intersection of these strips. Proving that h is continuous and that h conjugates f follows in the same way as the arguments of Theorem 5.2.3. □

7.2. Restricted Horseshoes

Consider $\widehat{R}_1 = [0, 1/3] \times [2/3, 1] \subset [0, 1] \times [2/3, 1] = R_1$, and $\widehat{R}_0 = R_0$. Then consider $g = f|_{\widehat{R}_0 \times \widehat{R}_1}$. Let us investigate the structure of Λ_g. Note that $(0, 0) \in \Lambda_g$ since $g(0, 0) = f(0, 0) = (0, 0)$ so every forward and backward iterate of $(0, 0)$ lies in the domain of g. We could attempt a geometric argument to show that Λ_g is interesting an nontrivial (ie, a Cantor set), but let us try to use the tools of Markov partitions. We wish to define again:

$$h(\mathbf{x}) = \bigcap_{i\in\mathbb{Z}} g^{-i}(\widehat{R}_{x_i})$$

Sadly, this will not work for us! For instance, consider the sequence $\mathbf{1} = (1, 1, 1, 1, 1, \dots)$. Note that for g, however, $g(\widehat{R}_1) \cap \widehat{R_1} = \emptyset$. Thus, we cannot define h this way. In fact, since $g(\widehat{R}_1) \cap \widehat{R_1} = \emptyset$, any sequence with 2 consecutive 1's will fail to have an well-defined $h(\mathbf{x})$. Thus we need yet another shift space:

DEFINITION 7.2.1. Let $A = (a_{ij})$ be a $d \times d$ matrix with entriess either 0 or 1. Define $\Sigma_A = \{\mathbf{x} \in \Sigma_d : A_{x_k x_{k+1}} = 1 \; \forall k \in \mathbb{Z}\}$. Σ_A is called a *subshift of finite type*.

Note that Σ_A is shift invariant (ie, $\sigma(\Sigma_A) = \Sigma_A$), and if it inhertis the ambient metric, is compact. To see how this relates to the restricted Horseshoe, construct the matrix by setting $a_{ij} = 1$ if $g(\widehat{R}_i) \cap \widehat{R}_j \neq \emptyset$. In this case, our matrix A is:

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

PROPOSITION 7.2.2. *For the map $g : \Lambda_g \to \Lambda_g$, h is well defined on Σ_A, and conjugates g with the shift $\sigma : \Sigma_A \to \Sigma_A$.*

SKETCH OF PROOF. Again, we only provide a sketch. Our arguments of continuity and conjugacy holds if we can show that h is well-defined and bijective. We thus need to show that for each sequence $\mathbf{x} \in \Sigma_A$, there exists a unique point $y \in \Lambda_g$ such that $g^i(y) \in \widehat{R}_{x_i}$. It is obvious that if such a point exists, its symbolic orbit must lie in Σ_A. We leave it to the reader to check that this is well defined for every such sequence [*Hint*: Again isolate the variables, and note that the local inverses of the maps are no longer onto!] □

We remark that the theory of Markov partitions holds in a much greater generality, but that to introduce them in such generality would require many technical definitions and tiresome work. We content ourselves with this special case, and direct interested parties to the book of Katok-Hasselblatt, Section 18.7.

7.3. Toral Endomorphisms

Let us return to the natural geometric notions in $\mathbb{R}$ and $\mathbb{Z}$. Recall that we considered the covering $\mathbb{R} \to \mathbb{R}/\mathbb{Z}$ and that $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ is homeomorphic to S^1 (in fact, isomorphic as groups). Then we can define $\mathbb{T}^d \cong \mathbb{R}^d/\mathbb{Z}^d \cong (\mathbb{R}/\mathbb{Z})^d$ ($\mathbb{T}^d$ is called the *torus of dimension* d), and are interested in studying the topological and dynamical properties of this space.

PROPOSITION 7.3.1. *If $p_1 : X_1 \to Y_1$ and $p_2 : X_2 \to Y_2$ are covering maps, $p_1 \times p_2 : X_1 \times X_2 \to Y_1 \times Y_2$ is a covering map*

PROOF. Let $\mathcal{U} = \{U_\alpha : \alpha \in A\}$ and $\mathcal{V} = \{V_\beta : \beta \in B\}$ be the open covers of Y_1 and Y_2, respectively, which satisfy the covering property. Then one can easily check from properties of product spaces that $\mathcal{W} = \{U_\alpha \times V_\beta : \alpha \in A, \beta \in B\}$ satisfyies the covering property for $p_1 \times p_2$. □

COROLLARY 7.3.2. *$\pi : \mathbb{R}^d \to \mathbb{T}^d$ sending $\pi(x_1, x_2, \dots, x_d) = ([x_1], \dots, [x_d])$ is a covering map.*

We now state the analog of the degree and lifting theorems that appeared in Section 4.1:

PROPOSITION 7.3.3. *If $f : \mathbb{T}^d \to \mathbb{T}^d$ is a continuous map, there exists a lift $\tilde{f} : \mathbb{R}^2 \to \mathbb{R}^2$ such that $\pi \circ \tilde{f} = f \circ \pi$ (where π is the canonical projecction), and any lift takes the form $\tilde{f}'(x) = \tilde{f}(x) + k$ for some $k \in \mathbb{Z}^d$. Furthermore, there exists a unique $\mathbb{Z}^2$-periodic function u (ie, $u(t) = u(t + k)$ for $k \in \mathbb{Z}^d$) and a unique $d \times d$ matrix with entries in $\mathbb{Z}$ such that:*

$$\tilde{f}(x) = Ax + u(x)$$

PROOF. Let f be a continuous map as described, and $\pi_k : \mathbb{T}^d \to \mathbb{T}$ be the projection onto the k^{th} coordinate (that is, $\pi_k([x_1], \ldots, [x_d]) = [x_k]$). Note that for each $x \in \mathbb{T}^d$ we can choose a path $c_x : [0,1] \to \mathbb{T}^d$ by setting $c_x(t) = [tx]$. Then if we set $c_{k,x} = \pi_k \circ c_x$, by the path lifting property for $\mathbb{T}$ (Theorem 4.1.1), we can choose a unique path $\tilde{c}_{k,x} : [0,1] \to \mathbb{R}$ such that $p \circ \tilde{c}_{k,x} = c_{k,x} \circ p = \pi_k \circ c_x \circ p$ by specifying a fixed start point. Then define the path $\tilde{c}_x : [0,1] \to \mathbb{R}^d$ by setting $\tilde{c}_x(t) = (c_{1,x}(t), \ldots, c_{d,x}(t))$. Then $\pi \circ c_x = c_x \circ \pi$ (we have thus proven the path lifting property for Tori).

In the same way as we proved the existence of lifts, here again we can define $\tilde{f}(x) = \tilde{c}_{f(x)}(1)$, and the proof here changes nowhere. Let us now prove the analog of degree. Again, any two lifts of a function must differ by a constant in $\mathbb{Z}^d$. That is, if $\tilde{f}_1$ and $\tilde{f}_2$ are two lifts of the same function $f : \mathbb{T}^d \to \mathbb{T}^d$, then $\tilde{f}_1 - \tilde{f}_2 \equiv (k_1, \ldots, k_d)$ (one can check this in the same way as before, observing that $\mathbb{Z}^d$ is discrete). Furthermore, for any given lift $\tilde{f}$, $\tilde{f}(x+k) = \tilde{f}(x) + \varphi_x(k)$, where $\varphi_x : \mathbb{Z}^d \to \mathbb{Z}^d$ is some map (k represents a d-tuple of integers). We claim that φ_x is independent of x and in fact a homomorphism. Let us first observe that it is independent of x (the homomorphism property will follow easily from this). Note that the function $x \mapsto \tilde{f}(x+k)$ is again a lift of $\tilde{f}$, so there exists some fixed $l \in \mathbb{Z}^d$ such that $\tilde{f}(x+k) = \tilde{f}(x) + l$ for every $x \in \mathbb{R}^d$. But then $\varphi_x(k) = l = \varphi_y(k)$, since l is independent of x.

So we are justified in writing φ. Finally, note that:

$$\varphi(k+l) = \tilde{f}(x+k+l) - \tilde{f}(x) = \tilde{f}(x+k+l) - \tilde{f}(x+l) + \tilde{f}(x+l) - \tilde{f}(x) = \varphi(k) + \varphi(l)$$

Observe also that if φ extended to a linear map $A : \mathbb{R}^d \to \mathbb{R}^d$, we could define $u(x) = \tilde{f}(x) - \varphi(x)$ (see Lemma 7.3.4). We claim then that u is $\mathbb{Z}^d$-periodic, since:

$$u(x+k) = \tilde{f}(x+k) - \varphi(x+k) = \tilde{f}(x) + \varphi(k) - \varphi(x) - \varphi(k) = \tilde{f}(x) - \varphi(x) = u(x)$$

This completes the proof of existence. We refer the reader again to Section 4.1 for a proof of uniqueness. □

LEMMA 7.3.4. *If $\varphi : \mathbb{Z}^d \to \mathbb{Z}^d$ is a homomorphism, there exists a unique linear map $A : \mathbb{R}^d \to \mathbb{R}^d$ such that $A|_{\mathbb{Z}^d} = \varphi$. Written as a matrix with the standard coordinates, A will have integer entries. A will be invertible with A^{-1} having integer entries if and only if* $\det(A) = \pm 1$.

PROOF. Let $e_1, \ldots, e_d$ represent the d basis vectors for $\mathbb{Z}^d$. That is, $e_1 = (1, 0, 0, \ldots, 0)$, $e_2 = (0, 1, 0, \ldots, 0)$, and so on. Note that each $k \in \mathbb{Z}^d$ can be expressed uniquely as $k = \sum_{i=1}^d k_i e_i$. Define $a_{ij} = (\varphi(e_i))_j$. That is, the j^{th} entry of the image of the i^{th} basis vector. Then the matrix A with entries a_{ij} has entries in the integers, and thus induces a linear transformation on $\mathbb{R}^d$. Observe that:

$$Ae_i = \sum_{j=1}^d a_{ij} e_j = \sum_{j=1}^d (\varphi(e_i))_j e_j = \varphi(e_i)$$

So $A|_{\mathbb{Z}^d} = \varphi$. Furthermore, A is unique, since it is determined by the matrix coefficiently a_{ij}. If one of them differed, it would induce a different map $A|_{\mathbb{Z}^d}$.

Finally, let us deal with invertibility. If A^{-1} exists and has integer entries, the $\det(A), \det(A^{-1}) \in \mathbb{Z}$, since the determinant is polynomial in the entries of the matrix. Thus, $1 = \det(\mathrm{id}) = \det(AA^{-1}) = \det(A)\det(A^{-1})$. The only integers which can multiply to 1 are ± 1, so in this case, $\det(A) = \det(A^{-1}) = \pm 1$.

To see the other direction, suppose that $\det(A) = \pm 1$. Cramer's Rule expresses the inverse of a matrix using the determinant and adjugate matrix: $A^{-1} = \det(A)^{-1} \cdot \mathrm{Adj}(A)$. Here, Adj represents an operation which only uses determinants of minors of the original matrix A. Thus, $\mathrm{Adj}(A)$ has integer entries, and since $\det(A) = \pm 1$, A^{-1} will have integer entries. □

Corollary 7.3.5. *If* $f : \mathbb{T}^d \to \mathbb{T}^d$ *is invertible, then the associated matrix is in* $GL(d, \mathbb{Z}) = \{A \in GL(n, \mathbb{R}) : A$ *has integer entries and* $\det(A) = \pm 1\}$*. Moreover, we can write:*

$$\tilde{f}^{-1}(x) = A^{-1}\left(x - u \circ \tilde{f}^{-1}\right)$$

We also make note of the convenient subgroup: $SL(d, \mathbb{Z}) = \{A \in GL(d, \mathbb{Z}) : \det(A) = 1\}$. Like in the case of $\mathbb{T}$, this is only a necessarily condition: there are maps which are not invertible, but still have their associated matrix in $GL(d, \mathbb{Z})$. The matrix A appears because we need to account for many more homomorphisms $\varphi : \mathbb{T}^2 \to \mathbb{T}^2$ appearing (note that in the case of $\mathbb{T}$, the only continuous self-homomorphisms are the maps L_d). For instance we could take the map:

$$f(x, y) = (2x + \sin(2\pi x), 3y + \cos(2\pi(x + y)))$$

Then the matrix A in this case would be $\begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}$. Other fundamental examples include the following maps:

$$\begin{aligned} (x, y) &\mapsto (2x + y, x + y) \\ (x, y) &\mapsto (x + \alpha, y + \beta) \\ (x, y) &\mapsto (x + \alpha, y + x) \end{aligned}$$

Note that all of the systems above are invertible. This is in stark contrast the the case of $\mathbb{T}$, when there were only two feasible degrees to obtain a homomorphism. These three maps on $\mathbb{T}^2$ are the starting point for many areas of study in dynamics, and will be our focus for much of the rest of the course!

CHAPTER 8

Surface Dynamics

In this Chapter, we will discuss homeomorphisms, $f : S \to S$, when S is a *surface*. We will not give the rigorous definition, but satisfy ourselves with some examples, such at $\mathbb{T}^2$. Other natural spaces that may appear are *cylinders*. This includes $S^1 \times \mathbb{R} = \mathbb{R}/\mathbb{Z} \times \mathbb{R} = \mathbb{R}^2/\mathbb{Z}$. We can also consider $S^1 \times [0,1] = \mathbb{R}/\mathbb{Z} \times [0,1]$. Another fundamental surface is the 2-sphere:

$$S^2 = \left\{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = 1\right\}$$

In the next section, we will see how to construct many surfaces starting from just $\mathbb{R}^2$ and S^2.

8.1. Examples of Surfaces

As discussed above, many surfaces are obtained as quotients of $\mathbb{R}^2$. This process is very general, we present a brief discussion here.

Definition 8.1.1. If X is a metric space, $f : X \to X$ is an *isometry* if $d(f(x), f(y)) = d(x, y)$. The *isometry group* of X is the group $\mathrm{Isom}(X)$, the set of bijective isometries of X.

Note that every $f \in \mathrm{Isom}(X)$ is a homeomorphism (since we insisted it is bijective). We can put a topology on $\mathrm{Isom}(X)$ by letting it inherit the sup-norm. That is, if $f, g \in \mathrm{Isom}(X)$, set:

$$\hat{d}(f, g) = \sup_{x \in X} \{d(f(x), g(x)\}$$

The subgroup Γ is said to act *properly discontinuously* if there exists a $\varepsilon > 0$ such that whenever there is some $x \in X$ and $\gamma \in \Gamma$ satsifying $d(x, \gamma(x)) < \varepsilon$, we know that $\gamma = \mathrm{id}$. Every subgroup group that acts properly discontinuously is discrete, since if $\gamma' \in B_\varepsilon(\gamma)$, $d(\gamma'(x), \gamma(x)) = d(\gamma^{-1}(\gamma'(x)), x) < \varepsilon$ for every x. Hence $\gamma' = \gamma$.

The converse is not true. For instance take the group of rotations around the origin which are multiples of $\pi/2$ (it is a group with 4 elements). Every element of the group fixes the origin, hence $d(\gamma(0), 0) = 0$ for every $\gamma \in \Gamma$. Hence this group does not act properly discontinously. This property can be thought of as taking a local property (the fact that γ moves some point) to a global property (the fact that γ moves every point).

Let $\Gamma \subset \mathrm{Isom}(X)$ be a discrete subgroup of $\mathrm{Isom}(X)$. Then define X/Γ to be the *orbit space*. That is, $X/\Gamma = \{\Gamma \cdot x : x \in X\}$, where $\Gamma \cdot x = \{\gamma(x) : \gamma \in \Gamma\}$. Points in X/Γ correspond to orbits. For example, if $X = \mathbb{R}$, and $\Gamma = \{T_n : n \in \mathbb{Z}\}$, then $X/\Gamma = \mathbb{R}/\mathbb{Z}$, since $\Gamma \cdot x = x + \mathbb{Z}$.

Proposition 8.1.2. *Let X be a metric space. Then if Γ is a group of isometries acting properly discontinuously on X, X/Γ is a metric space, with the metric:*

$$\bar{d}(\Gamma \cdot x, \Gamma \cdot y) = \inf_{\gamma \in \Gamma} \{d(x, \gamma(y))\}$$

Furthermore, the projection $p : X \to X/\Gamma$ mapping $x \mapsto \Gamma \cdot x$ is a covering map.

Proof. Let us first show that $\bar{d}$ is a metric. First, we'll show symmetry. Note that if $\bar{d}(\Gamma \cdot x, \Gamma \cdot y) = r$, then there are a sequence of isometries γ_n such that $d(x, \gamma_n(y)) \to r$. But since γ_n is an isometry, $d(y, \gamma_n^{-1}(x)) = d(\gamma_n^{-1}(x), y) = d(x, \gamma_n(y)) \to r$. Thus $\bar{d}(\Gamma \cdot y, \Gamma \cdot x) \le \bar{d}(\Gamma \cdot x, \Gamma \cdot y)$. Symmetry follows by switching the roles of x and y.

Let us now check the triangle inequality. This follows since if $\bar{d}(\Gamma\cdot x, \Gamma\cdot y) = r$ and $\bar{d}(\Gamma\cdot y, \Gamma\cdot z) = s$, there exist sequences γ_n and δ_n such that $d(x, \gamma_n(y)) \to r$ and $d(y, \delta_n(z)) \to s$. But then:

$$d(x, \gamma_n(\delta_n(z))) \leq d(x, \gamma_n(y)) + d(\gamma_n(y), \gamma_n(\delta_n(z))) = d(x, \gamma_n(y)) + d(y, \delta_n(z)) \to r + s$$

Thus, $\bar{d}(\Gamma \cdot x, \Gamma \cdot z) \leq r + s = \bar{d}(\Gamma \cdot x, \Gamma \cdot y) + \bar{d}(\Gamma \cdot y, \Gamma \cdot z)$.

Finally, we need to check that $\bar{d}(\Gamma \cdot x, \Gamma \cdot y) = 0$, then $\Gamma \cdot x = \Gamma \cdot y$. Suppose that we have $\bar{d}(\Gamma \cdot x, \Gamma \cdot y) = 0$, so that there are a sequence of isometries γ_n such that $d(\gamma_n(x), y) \to 0$. Thus, $\gamma_n(x)$ is a Cauchy sequence, and for some N if $n \geq N$, $d(\gamma_n^{-1}\gamma_N(x), x) < \varepsilon$ (here ε is as in the definition of acting properly discontinuously). Thus, $\gamma_n^{-1}\gamma_N = \mathrm{id}$ and $\gamma_n = \gamma_N$ for $n \geq N$. This implies that $y = \gamma_N(x)$ and $\Gamma \cdot y = \Gamma \cdot x$.

To see that the projection is a covering map, we choose the open cover of X/G to be $\mathcal{U} = \{B_\varepsilon([x]) : [x] \in X/G\}$, where $2\varepsilon > 0$ is as in the proper discontinuity of Γ. Then, for each $B_\varepsilon([x])$, $p^{-1}(B_\varepsilon([x])) = \bigcup_{\gamma\in\Gamma} B_\varepsilon(\gamma(x))$. But, if $B_\varepsilon(\gamma_1(x)) \cap B_\varepsilon(\gamma_2(x)) \neq \emptyset$, $B_\varepsilon(x) \cap B_\varepsilon(\gamma_1^{-1}\gamma_2(x)) \neq \emptyset$. But then $d(x, \gamma_1^{-1}\gamma_2(x)) < 2\varepsilon$, and $\gamma_1 = \gamma_2$. Thus, we indeed have that $p^{-1}(B_\varepsilon([x])) = \bigsqcup_{\gamma\in\Gamma} B_\varepsilon(\gamma(x))$. □

Using Proposition 8.1.2 we can introduce some prototypical surfaces. First, we observe that the Torus can be obtained this way using the group $\Gamma = \left\{T_n : n \in \mathbb{Z}^d\right\}$, where $T_v : \mathbb{R}^d \to \mathbb{R}^d$ is the map $T_v(x) = x + v$.

The Möbius band can bet obtained from $\mathbb{R}^2$ by using the map $F(x, y) = (x + 1, -y)$, so that $F^n(x, y) = (x + n, (-1)^n y)$. One easily observes that the group $\Gamma = \{F^n : n \in \mathbb{Z}\}$ is a group of isometries and acts properly discontinuously. The quotient $\mathbb{R}^2/\Gamma$ is called the Möbius Band. It looks as if it were a cylinder. Each point in the quotient has a unique representative in $[0, 1) \times \mathbb{R}$. If we add in $\{1\} \times \mathbb{R}$, we are forced to identify the points $(0, t)$ with $(1, -t)$.

It is not immediately obvious that the Möbius band and the cylinder are not homeomorphic, but they are in fact distinct topological spaces. One easy way to see this is that every loop in the cylinder divides it into two connected components, but there exists a loop in the Möbius band for which the resulting space is still connected. This will not be proven and we leave it to the reader to find such loops.

We will list some other groups Γ and spaces X to get some surfaces:

Sphere	S^2	$\Gamma = \{e\}$	$X = S^2$
Plane	$\mathbb{R}^2$	$\Gamma = \{e\}$	$X = \mathbb{R}^2$
Projective Plane	$\mathbb{R}P^2$	$\Gamma = \{\pm\,\mathrm{id}\}$	$X = S^2$
Torus	$\mathbb{T}^2$	$\Gamma = \left\{T_n : n \in \mathbb{Z}^2\right\}$	$X = \mathbb{R}^2$
Möbius Band	M	$\Gamma = \{F^n : n \in \mathbb{Z}\}$	$X = \mathbb{R}^2$
Klein Bottle	K	$\Gamma = \left\{T_{(0,m)} \circ F^n : m, n \in \mathbb{Z}\right\}$	$X = \mathbb{R}^2$

See Figures 1-3.

8.2. Fixed Points of Endomorphisms

Recall from Proposition 7.3.3 that any homeomorphism $f : \mathbb{T}^d \to \mathbb{T}^d$ can be lifted to a homeomorphism $\tilde{f} : \mathbb{R}^d \to \mathbb{R}^d$ of the form $\tilde{f} = A + u$, where u is $\mathbb{Z}^d$-periodic and $A \in GL(d, \mathbb{Z})$. This mostivates to study the prototypical case when $u = 0$. In this case, we often abuse notation and just write $f = A : \mathbb{T}^d \to \mathbb{T}^d$, so that A is a continuous homomorphism of both $\mathbb{T}^d$ and $\mathbb{R}^d$.

Our first seemingly simple task is this: How can we detect fixed points of A? In the case of $d = 1$, the only possible maps were L_n and we discovered that $|\mathrm{Fix}(L_n)| = n - 1$, and the fixed points corresponded to the set $F = \{k/(n-1) : k \in \mathbb{Z}\} \pmod 1$. In the case when $d > 1$, we need to solve the following equation for some $n \in \mathbb{Z}^d$:

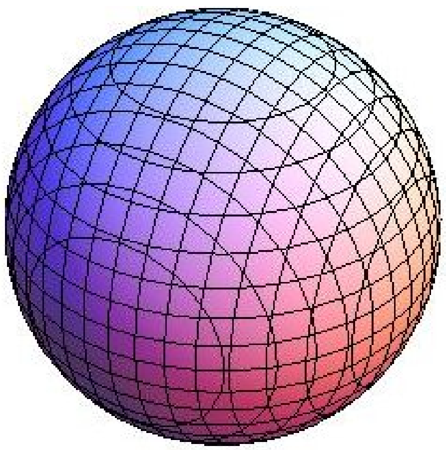

FIGURE 1. The 2-sphere

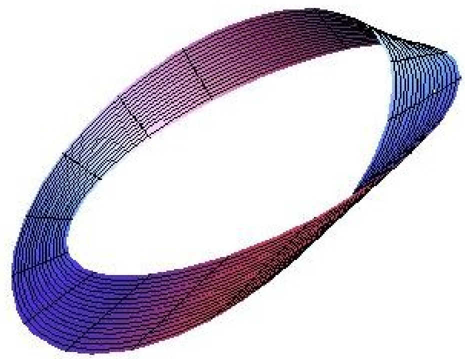

FIGURE 2. The Möbius Band as a subset of $\mathbb{R}^3$

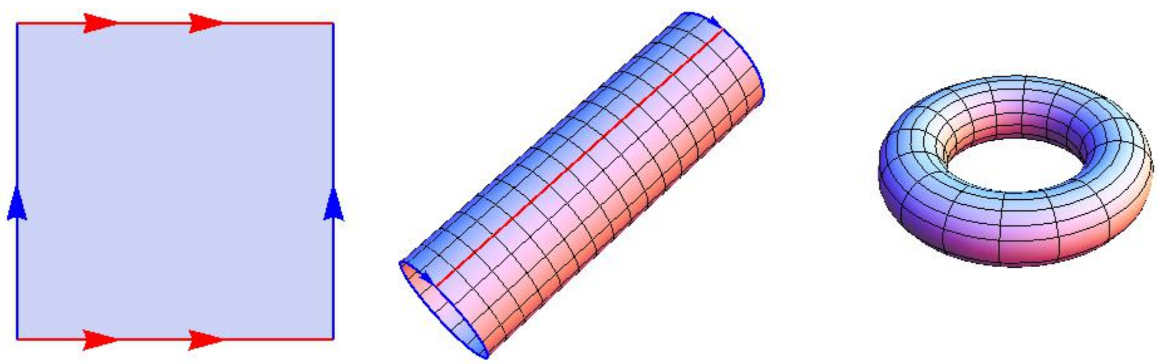

FIGURE 3. The torus obtained as quotient spaces

$$\begin{aligned} Ax &= x+n \\ (A-\mathrm{id})x &= n \\ x &= (A-\mathrm{id})^{-1}n \end{aligned}$$

This computation works, assuming that $A - \mathrm{id}$ is invertible (ie, that 1 is not an eigenvalue of A). If 1 is an eigenvalue, then the eigenspace corresponding to 1 is fixed, so $|\mathrm{Fix}(A)| = \infty$. Thus we have shown that:

PROPOSITION 8.2.1. *If* $\det(A-\mathrm{id}) \neq 0$*, then* $\mathrm{Fix}(A) = (A-\mathrm{id})^{-1}(\mathbb{Z}^d) \pmod 1$

Counting the fixed points will be slightly more difficult. We need to find a way of expressing $F = (A-\mathrm{id})^{-1}(\mathbb{Z}^d) \cap [0,1)^d$, since each point of $\mathbb{T}^d$ has exactly one representative in $[0,1)^d$.

PROPOSITION 8.2.2. *If* $A : \mathbb{T}^d \to \mathbb{T}^d$ *has* $\det(A-\mathrm{id}) \neq 0$*, then* $|\mathrm{Fix}(A)| \leq |\det(A-\mathrm{id})|^d$

PROOF. We continue using the notation $F = (A-\mathrm{id})^{-1}(\mathbb{Z}^d) \cap [0,1)^d$. We first claim that $|F| < \infty$. Note that $(A-\mathrm{id})^{-1} = \det(A-\mathrm{id})^{-1}B$ for some integer matrix B (we use Cramer's rule as in the proof of Lemma 7.3.4). Thus,

$$(A-\mathrm{id})^{-1}(\mathbb{Z}^d) = \det(A-\mathrm{id})^{-1}B(\mathbb{Z}^d) \subset \det(A-\mathrm{id})^{-1}(\mathbb{Z}^d)$$

Thus, the fixed points are contained in the set of rational d-tuples that can be written with denominator $\det(A-\mathrm{id})$. In particular, $|F| \leq \det(A-\mathrm{id})^d$. □

This inquality is very bad unless $d = 1$. In fact, we will show that $|\mathrm{Fix}(A)| = |\det(A-\mathrm{id})|$, but this will take more work than one might expect, and will consume our discussion for the rest of this section.

We know by Proposition 8.2.1 that $\mathrm{Fix}(A) = (A-\mathrm{id})^{-1}(\mathbb{Z}^d) \pmod 1$ (here $\pmod 1$ means projecting to the torus from $\mathbb{R}^d$). Thus, if $B = A - \mathrm{id}$, then $x \in \mathrm{Fix}(A)$ if and only if $B(x) = [0]$, so we would like to study $B^{-1}([0])$. In particular, we would like to study $|\mathrm{Fix}(A)| = \left|B^{-1}(0) \cap [0,1)^d\right|$.

Proposition 8.2.3. *If $B : \mathbb{T}^d \to \mathbb{T}^d$ is a toral endomorphism and $\det B \neq 0$, then B is onto and $\left|p(B^{-1}(\mathbb{Z}^d))\right| = \left|p(B^{-1}(x + \mathbb{Z}^d))\right|$, where $x \in \mathbb{R}^d$ and $p : \mathbb{R}^d \to \mathbb{T}^d$ is the projection*

Proof. The fact that $\det B \neq 0$ implies that $B : \mathbb{R}^d \to \mathbb{R}^d$ is onto. Thus, $[0,1)^d \subset B(\mathbb{R}^n)$, and every point in $[0,1)$ has at least one preimage in $\mathbb{R}^d$, and projecting this preimage to $\mathbb{T}^d$, we see that $B : \mathbb{T}^d \to \mathbb{T}^d$ is onto.

If $\{x_1, \dots, x_r\} = p(B^{-1}(\mathbb{Z}^d))$ and $B(y) = x$, then $B(y + x_i) \in x + \mathbb{Z}^d$, since B is linear. So $\left|p(B^{-1}(x + \mathbb{Z}^d))\right| \geq \left|p(B^{-1}(\mathbb{Z}^d))\right|$. Switching the roles of x and 0 gives the result. □

Definition 8.2.4. If B is a toral endomorphism, let $\deg(B) = \mathrm{sgn}(\det B) \cdot \left|B^{-1}(0)\right|$ if $\det(B) \neq 0$ and $\deg(B) = 0$ if $\det(B) = 0$

Note that in the case of $\mathbb{R}/\mathbb{Z}$ this coincides with the degree we introduced before: $\deg(L_d) = \left|L_d^{-1}(0)\right| = |\{k/d : k = 0, \dots, d-1\}| = d$.

Proposition 8.2.5. $\deg(AB) = \deg(A)\deg(B)$

Proof. If any of A, B or AB is degenerate, then one easily observes that the formula holds. So let us assume that $\det(A), \det(B) \neq 0$. We wish to show that:

$$\left|(AB)^{-1}(0)\right| = \left|A^{-1}(0)\right| \cdot \left|B^{-1}(0)\right|$$

(We have dropped the projection p, it is assumed that the preimages are taken on $\mathbb{T}^d$). Note that $(AB)^{-1}(0) = B^{-1}A^{-1}(0)$, so if $A^{-1} = \{x_1, \dots, x_r\}$, then $(AB)^{-1}(0) = B^{-1}(\{x_1, \dots, x_r\}) = B^{-1}(x_1) \cup B^{-1}(x_2) \cup \dots B^{-1}(x_r)$. Note that each of these sets are disjoint and has the same cardinality by Proposition 8.2.3, $\left|(AB)^{-1}(0)\right| = \sum_{i=1}^r \left|B^{-1}(0)\right| = \left|A^{-1}(0)\right| \left|B^{-1}(0)\right|$. □

The determinant clearly plays an important role, so let us find a geometric understanding of the determinant:

Proposition 8.2.6. *Let $A : \mathbb{R}^d \to \mathbb{R}^d$ be a linear map. Then* $\mathrm{vol}(A(I^d)) = |\det(A)| \,\mathrm{vol}(I^d) = |\det(A)|$

Proof. The proof takes two steps. First we show it for elementary matrices: diagonal matrices, and upper- and lower-triangular matrices of the form $E_{ij}(\lambda) = \mathrm{id} + \lambda e_{ij}$, where e_{ij} is the matrix with entries $(e_{ij})_{kl} = \delta_{ik}\delta_{jl}$. Then we show that if the formula holds for two matrices A, B then it holds for their product. Note that diagonal matrices are trivial, since cubes will be taken to cubes whose side have lengths exactly the entries of the diagonal matrix. Thus, its volume is the product of these lengths, which is exactly the determinant.

We will show that this holds for the matrices $E_{ij}(\lambda)$ only in dimension 2. Note that the cube is transformed into a parallelogram with the same area (see Figure 4). The argument holds in higher dimensions using parallelepipeds, observing that if you look at the $(d-1)$-dimensional space which is fixed by $E_{ij}(\lambda)$ (namely $V = \mathbb{R} \times \mathbb{R} \times \cdots \times \{0\} \times \cdots \times \mathbb{R}$, where we put the $\{0\}$ is the j^{th} slot).

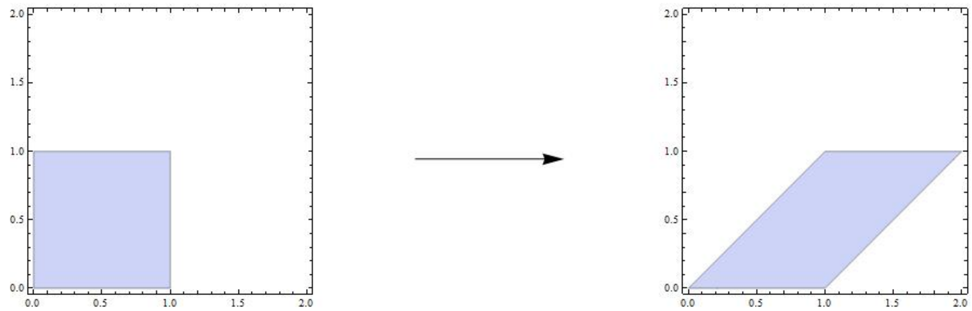

FIGURE 4. The matrix $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ acting on I^2

Then $E_{ij}(\lambda)$ translates each V-slice by a fixed amount, so if we express the volume as the integral over the $(d-1)$-dimensional slices and note that translation will preserve the $(d-1)$-dimensional volume, $E_{ij}(\lambda)$ must preserve the volume of the cube. $\square$

PROPOSITION 8.2.7. *If* $B : \mathbb{T}^d \to \mathbb{T}^d$ *is a toral endomorphism, then* $\deg(B) = \det(B)$

Before proving the proposition, let us observe that:

COROLLARY 8.2.8. *If* $A : \mathbb{T}^d \to \mathbb{T}^d$ *is a toral endomorphism with* $\det(A - \mathrm{id}) \neq 0$*, then* $|\mathrm{Fix}(A)| = |\det(A - \mathrm{id})|$*, and hence if* $\det(A^n - \mathrm{id}) \neq 0$ *then* $|\mathrm{Fix}(A^n)| = |\det(A^n - \mathrm{id})|$

PROOF OF PROPOSITION 8.2.7. Note that $\left|B^{-1}(0)\right| = \left|B^{-1}(\mathbb{Z}^d) \cap R\right|$, where $R = [-1/2, 1/2)^d$. Furthermore, $\left|B^{-1}(\mathbb{Z}^d) \cap R\right| = \left|\mathbb{Z}^d \cap B(R)\right|$ (since B is invertible on $\mathbb{R}^d$ it preserves cardinalities and distributes into intersections). Our first step is to show that:

$$2^{-d} \left|\det(B)\right| \leq \left|B^{-1}(0)\right| \leq 2^d \left|\det(B)\right| \tag{5}$$

Let us first prove it in the case when $R \subset B(R)$ (note that this is a nontrivial assumption, and will be resolved later!). Suppose that $B(R) \cap \mathbb{Z}^d = \{n_1, \ldots, n_r\} \subset \mathbb{Z}^d$. We claim that for each n_i, $n_i + R \subset B(R) + R \subset B(R) + B(R) = 2B(R)$. Here, we use the conventions that $X + Y = \{x + y : x \in X, y \in Y\}$ and $2X = \{2x : x \in X\}$. These conventions make the equation $X + X = 2X$ a nontrivial fact (and indeed false if X is not convex). Nonetheless, it holds for $X = B(R)$ since $B(R)$ is convex. It is trivial that $2B(R) \subset B(R) + B(R)$. Suppose that $x + y \in B(R) + B(R)$. Since R is convex, $B(R)$ is convex and $z = (x + y)/2 \in B(R)$. In particular, $x + y = 2z \in 2B(R)$ and $B(R) + B(R) \subset 2B(R)$.

Now, $\bigcup_{i=1}^r n_i + R \subset 2B(R)$, so:

$$\mathrm{vol}\left(\bigcup_{i=1}^r n_i + R\right) \leq \mathrm{vol}(2B(R)) = |\det(2B)| = 2^d \left|\det(B)\right|$$

Thus, $\left|B^{-1}(0)\right| \leq 2^d \left|\det(B)\right|$.

Now we seek a bound from below. Let $x \in B(R)$, and $n \in \mathbb{Z}^d$ be such that $x \in n + R$ (such and n always exists since each $x \in \mathbb{T}^d$ has a unique representative in R). Thus, $n \in x - R \subset B(R) - R \subset B(R) + B(R) = 2B(R)$. Then $B(R) \subset \bigcup_{n \in \mathbb{Z}^d \cap 2B(R)} n + R$. In particular,

$$|\det(B)| = \mathrm{vol}(B(R)) \leq \left|\mathbb{Z}^d \cap 2B(R)\right| = |\deg(2B)| = |\deg(2\,\mathrm{id}) \cdot \deg(B)| = 2^d \left|\deg(B)\right|$$

That is, $|\deg(B)| \geq 2^{-d} \left|\det(B)\right|$. Combining the lower and upper bounds, we find that under the assumption that $R \subset B(R)$ we have exactly (5). (Observe that we used that $\deg(q\,\mathrm{id}) = q^d$ for $q \in \mathbb{Z}$, this is left as an exercise for the reader.)

Let us now remove our condition $R \subset B(R)$. Let B be any integer matrix such that $\det(B) \neq 0$. Thus, there exists $q \in \mathbb{N}$ such that $B(R) \supset \frac{1}{q}R$ (this follows from continuity of B). But then $qB(R) \supset R$, and if $\hat{B} = qB$, $\hat{B}(R) \supset R$. We get the bounds (5) for $\hat{B}$, but we know that $\deg(\hat{B}) = \deg(q\,\mathrm{id})\deg(B) = q^d \deg(B)$, and $\det(\hat{B}) = q^d \det(B)$. Cancelling the q^d term, we get (5) for B.

Now, note that for any $k \geq 0$:

$$2^{-d}\left|\det(B)\right|^k = 2^{-d}\left|\det(B^k)\right| \leq \left|\deg(B^k)\right| = \left|\deg(B)\right|^k \leq 2^d\left|\det(B^k)\right| = 2^d\left|\det(B)\right|^k$$

Taking the k^{th} root of each term, we find that:

$$2^{-d/k}\left|\det(B)\right| \leq \left|\deg(B)\right| \leq 2^{d/k}\left|\det(B)\right|$$

Letting $k \to \infty$, we get that $|\deg(B)| = |\det(B)|$. So $\deg(B)/\det(B) = \pm 1$. But since $\mathrm{sgn}(\deg(B)) = \mathrm{sgn}(\det(B))$, we get that $\deg(B) = \det(B)$. □

With a little algebra, we can find another proof. We present the proof for those familiar with some elementary abstract algebra and abelian group structure:

Definition 8.2.9. Let G be an abelian group, and H a subgroup. The *index of H in G* is defined to be $[G : H] = |G/H|$

Alternate Proof of Proposition 8.2.7. Note that the degree is exactly $\left|B^{-1}(0)\right|$ (here the inverse image is taken as a subset of $\mathbb{T}^d$). If $B(x) = 0$, $x \in \mathbb{T}^d$, then $\widetilde{B}(x) \in \mathbb{Z}^d$, where $\widetilde{B} : \mathbb{R}^d \to \mathbb{R}^d$ is the lift of B which is a homomorphism. We claim that the number of distinct points of $\mathbb{T}^d$ which are sent to 0 is exactly $[\mathbb{Z}^d : \widetilde{B}(\mathbb{Z}^d)]$. First, list $\ker B = B^{-1}(0) = \{[x_1], \dots, [x_n]\}$. Let us lift this to $\mathbb{R}^d$. Note that if $B([x_i]) = 0$, then $\widetilde{B}(x_i) \in \mathbb{Z}^d$. Furthermore, if $n \in \mathbb{Z}^d$, then $\widetilde{B}(x_i + n) \in \mathbb{Z}^d$, so that $\widetilde{B}(x_i + \mathbb{Z}^d) \in \mathbb{Z}^d$. Furthermore, if $m \in \mathbb{Z}^d$, then $y = \widetilde{B}^{-1}(m)$ has $\widetilde{B}(y) \in \mathbb{Z}^d$, so that $[y] \in \ker B$, and $[y] = [x_i]$ for some i, and $y \in x_i + \mathbb{Z}^d$. We have thus shown that $\mathbb{Z}^d = \bigcup_{i=1}^n B(x_i + \mathbb{Z}^d)$. We claim that this union is disjoint. Indeed, if $y \in \widetilde{B}(x_i + \mathbb{Z}^d) \cap \widetilde{B}(x_j + \mathbb{Z}^d)$, then $\widetilde{B}^{-1}y \in (x_i + \mathbb{Z}^d) \cap (x_j + \mathbb{Z}^d)$ and $[x_j] = [x_i]$. Thus, $i = j$, and the union must be disjoint. This shows that $|\deg(B)| = [\mathbb{Z}^d : B(\mathbb{Z}^d)]$.

So we wish to study $[\mathbb{Z}^d : B(\mathbb{Z}^d)]$. The plan will be similar to that of the geometric proof. We show that for certain, elementary generating matrices $|\det(A)| = [\mathbb{Z}^d : A(\mathbb{Z}^d)]$. Then, using Proposition 8.2.5, and the fact that $\det(AB) = \det(A)\det(B)$, we will get that $\det(A) = \deg(A)$. Let us assume the lemma and prove the formula for the elementary matrices. First, take a matrix of the form $A = \mathrm{diag}(a_1, \dots, a_d)$, representing a matrix with diagonal entries $a_i \in \mathbb{Z}$ and all other entries 0. Then $A(\mathbb{Z}^d) = (a_1\mathbb{Z}) \times (a_2\mathbb{Z}) \times \cdots \times (a_d\mathbb{Z})$, and $\mathbb{Z}^d/A(\mathbb{Z}^d) \cong (\mathbb{Z}/a_1\mathbb{Z}) \times \cdots \times (\mathbb{Z}/a_d\mathbb{Z})$. The group $\mathbb{Z}/a_i\mathbb{Z}$ will have $|a_i|$ elements, so their product will have $|a_1 a_2 \dots a_d| = |\det(A)|$ elements.

Let us consider the matrices $A = \mathrm{id} + ae_{ij}$ with $a \in \mathbb{Z}$, as in the proof of Proposition 8.2.6. We claim that $A(\mathbb{Z}^d) = \mathbb{Z}^d$ (and hence, that $[\mathbb{Z}^d : A(\mathbb{Z}^d)] = 1 = \det(A)$). To see this, note that $A^{-1} = \mathrm{id} - ae_{ij}$, so that if $m \in \mathbb{Z}^d$, then $m' = A^{-1}(m) \in \mathbb{Z}^d$ and $Am' = m$. Hence A is onto.

Each of the elementary matrices represents some elementary row operation in Gaussian elimination. Hence every matrix can be represented as a product of these elementary ones, when we allow entries in $\mathbb{Q}$. But if we have a finite product of matrices in $\mathbb{Q}$ $A = A_1 A_2 \dots A_N$, there exists some number M such that $MA = A_1' A_2' \dots A_N'$, where A_i' have integer entries (here, we are just clearing the denominators of A by multiplying by a large integer). But $MA = \mathrm{diag}(M, M, \dots, M)A$, so that $\det(MA) = M^d \det(A)$ and $\deg(MA) = M^d \deg(A)$, by Proposition 8.2.5. Thus, we have that $\deg(A) = \det(A)$. □

8.3. Fixed Points of Continuous Maps

We know that if $A : \mathbb{T}^d \to \mathbb{T}^d$ has finitely many fixed points, then $|\mathrm{Fix}(A)| = \left|\det(\widetilde{A} - \mathrm{id})\right|$, and the fixed points correspond exactly to the projections of the points $(\widetilde{A} - \mathrm{id})^{-1}(\mathbb{Z}^d)$. We wish to extend this to a bound on the number of fixed points in an arbitrary continuous map $f : \mathbb{T}^d \to \mathbb{T}^d$ of the form $f(x) = Ax + u(x)$. Indeed:

PROPOSITION 8.3.1. *If $f : \mathbb{T}^d \to \mathbb{T}^d$ takes the form $f(x) = Ax + u(x)$, then* $|\text{Fix}(f)| \geq |\det(A - \text{id})|$

Note that this is a direct generalization of Theorem 5.0.9.

PROOF. The proposition is vacuous when $\det(A - \text{id}) = 0$, so assume that $\det(A - \text{id}) \neq 0$. If for each fixed point of A, we can find distinct fixed points of f, then we prove the claim by applying Proposition 8.2.7. Observe that we can choose a finite number of integral points $n^{(i)} \in \mathbb{Z}^d$ such that the unique fixed points of the lifts $\widetilde{A}_i(x) = \widetilde{A}x + n^{(i)}$, y_i, represent the distinct fixed points of A, i.e. if $m \in \mathbb{Z}^d$ then $y_i + m \neq y_j$ if $i \neq j$. Note that we can choose $n^{(1)} = 0$, so that $\widetilde{A}(0) = 0$.

Fix a lift $\tilde{f} : \mathbb{R}^d \to \mathbb{R}^d$ of f. We seek a fixed point of $\tilde{f}$. This will occur only when $\widetilde{A}x + u(x) = x$, so we wish to solve $(\widetilde{A} - \text{id})x + u(x) = 0$, or $x = (\text{id} - A)^{-1}(u(x))$.

LEMMA 8.3.2. *If $u : \mathbb{R}^d \to \mathbb{R}^d$ is continuous and bounded, and $B : \mathbb{R}^d \to \mathbb{R}^d$ is a linear, invertible homomorphism, then:*

(1) $x \mapsto Bx + u(x)$ is onto
(2) There exists x_0 such that $B^{-1}(u(x_0)) = x_0$

PROOF. For each $y \in \mathbb{R}^d$, we wish to solve the equation $Bx + u(x) = y$, which is equivalent to solving $x + B^{-1}u(x) = B^{-1}y$, or $x = B^{-1}y - B^{-1}u(x)$. Then if $g : \mathbb{R}^d \to \mathbb{R}^d$ is defined by $g(x) = B^{-1}y - B^{-1}u(x)$, we seek a fixed point of g. Note that g is continuous and bounded, so if we can find the fixed point for an arbitrary bounded g, we're done.

To find a fixed point we shall use Brouwer's fixed point theorem (which we proved in the case of $d = 2$, Theorem 4.2.2). Hence we wish to find some $r > 0$ so that $g(\overline{B}_r(0)) \subset \overline{B}_r(0)$. In fact, we know that the image of g is contained in a closed ball since g is bounded, so this is not a problem. □

Let us state one last Lemma for clarity:

LEMMA 8.3.3. *Let $A : \mathbb{T}^d \to \mathbb{T}^d$ be a toral automorphism without 1 as an eigenvalue, and let $\widetilde{A}_1(x) = Ax + n_1$ and $\widetilde{A}_2(x) = Ax + n_2$ be two lifts of A (so that $n_i \in \mathbb{Z}^d$). Then if x_i is the unique fixed point of A_i in $\mathbb{R}^d$, x_1 and x_2 represent the same point on $\mathbb{T}^d$ if and only if $n_1 - n_2 \in (\text{id} - A)\mathbb{Z}^d$*

PROOF OF LEMMA 8.3.3. First suppose that $n_1 - n_2 \in (A - \text{id})\mathbb{Z}^d$. Then $x_1 - x_2 = Ax_1 + n_1 - (Ax_2 + n_2) = A(x_1 - x_2) + (n_1 - n_2)$, so $(\text{id} - A)(x_1 - x_2) = n_1 - n_2$. Since $\text{id} - A$ is invertible, we conclude that since $n_1 - n_2 \in (\text{id} - A)\mathbb{Z}^d$, then $x_1 - x_2 \in \mathbb{Z}^d$. That is, x_1 and x_2 represent the same point on $\mathbb{T}^d$.

Now suppose that x_1 and x_2 represent the same point on $\mathbb{T}^d$. Then $x_1 - x_2 \in \mathbb{Z}^d$, so that since $n_i = x_i - Ax_i$, $n_1 - n_2 = x_1 - Ax_1 - (x_2 - Ax_2) = (\text{id} - A)(x_1 - x_2) \in (\text{id} - A)\mathbb{Z}^d$. □

Choose fixed lifts $\tilde{f}_i(x) = \tilde{f}(x) + n_i = Ax + u(x) + n_i$ such that each n_i corresponds to a distinct fixed point of A (by Lemma 8.3.3, this is the largest collection of integral vectors such that if $n_i - n_j \in (\text{id} - A)^{-1}\mathbb{Z}^d$, then $i = j$). By Lemma 8.3.2, $\tilde{f}_i$ must also have a fixed point, so let x_i denote a fixed point of $\tilde{f}_i$. We wish to show that x_i are all distinct when projected to the torus.

Suppose that $x_i = x_j + m$ for some $m \in \mathbb{Z}^d$. Then observe that since $\tilde{f}_i(x_i) = x_i$, $x_j = Ax_j + u(x_j) + n_j$. Similarly:

$$x_i = \tilde{f}_i(x_i) = Ax_i + u(x_i) + n_i = A(x_j + m) + u(x_j + m) + n_i = Ax_j + u(x_j) + n_i + Am$$

Subtracting these equalities, we find:

$$m = x_i - x_j = n_i - n_j + Am$$

But then $n_i - n_j = (\mathrm{id} - A)m$, and since $m \in \mathbb{Z}^d$, we conclude from Lemma 8.3.3 that $i = j$ and the choices made for the n_i that $i = j$. That is, $m = 0$ and the x_i and x_j project to the same point if and only if $i = j$. □

Let us now return to the linear setting to pin down *exactly* what the periodic points are:

THEOREM 8.3.4. *Let $A : \mathbb{T}^d \to \mathbb{T}^d$ be a toral automorphism. Then:*

(i) $\mathrm{Per}(A) \supset \mathbb{Q}^d/\mathbb{Z}^d = \{[x] : x \in \mathbb{Q}^d\}$

(ii) If $\det(A^n - \mathrm{id}) \neq 0$ *for every* $n \geq 1$, *then* $\mathrm{Per}(A) = \mathbb{Q}^d/\mathbb{Z}^d$

PROOF. Let us first prove the first claim. Note that each point of $\mathbb{Q}^d/\mathbb{Z}^d$ has a unique representative in $\mathbb{Q}^d \cap [0,1)^d$. Note that we can write $\mathbb{Q}^d = \bigcup_{q\in\mathbb{N}} \frac{1}{q}\mathbb{Z}^d$ (this union is not disjoint, but this does not matter for our purposes). Thus, we can write that $\mathbb{Q}^d/\mathbb{Z}^d = \bigcup_{q\in\mathbb{N}} \left(\frac{1}{q}\mathbb{Z}^d\right)/\mathbb{Z}^d$. Each set in the union is a "net" in $\mathbb{T}^d$, which becomes finer and finer as $q \to \infty$. For instance, if $d = 2$, and $q = 4$, we get $16 = 4^2 = q^d$ points in the torus $\mathbb{T}^2$ which are placed at points with denominator 4 (see Figure 5, compare to Figure 4).

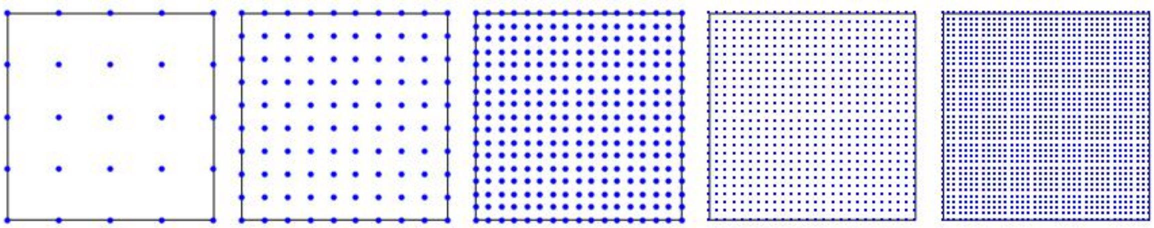

FIGURE 5. Subgroups of rational points in $\mathbb{T}^n$

LEMMA 8.3.5. *Let A be a toral automorphism, and $S_q = \left(\frac{1}{q}\mathbb{Z}^d\right)/\mathbb{Z}^d$.*

(a) $|S_q| = q^d$

(b) $A(S_q) = S_q$

PROOF. The first claim of the lemma follows from the fact that $\left(\frac{1}{q}\mathbb{Z}^d\right)/\mathbb{Z}^d \cong \left(\left(\frac{1}{q}\mathbb{Z}\right)/\mathbb{Z}\right)^d$. The second claim follows from the fact that if $\frac{1}{q}m \in \frac{1}{q}\mathbb{Z}^d$, then $\widetilde{A}\left(\frac{1}{q}m\right) = \frac{1}{q}(Am)$. But since A has integer entries, $Am \in \mathbb{Z}^d$, so $A(\frac{1}{q}\mathbb{Z}^d) \subset \frac{1}{q}\mathbb{Z}^d$. Since A is invertible, we get that $A(S_q) = S_q$. □

Now, since S_q is invariant for A, and S_q is finite, every $x \in S_q$ has two iterates of A which coincide (Take the first $q^d + 1$ iterates. All of them are in S_q, so by the pigeon hole principle, at least one appears twice). That is, $A^k x = A^l x$ for some $k \neq l$. Since A is invertible, this implies x is periodic.

Let us now prove the second claim. Again, we rely on a Lemma:

LEMMA 8.3.6. *If* $q = |\det(A - \mathrm{id})|$, *then* $\mathrm{Fix}(A) \subset \left(\frac{1}{q}\mathbb{Z}^d\right)/\mathbb{Z}^d$.

PROOF. If $B = A - \mathrm{id}$, the x is fixed if and only if $Bx = 0$, so $\mathrm{Fix}(A) = \widetilde{B}^{-1}(\mathbb{Z}^d)/\mathbb{Z}^d$ (see the discussion of the previous section). If $q = |\det(B)|$, then $B^{-1} = \frac{1}{q}C$, where C is a matrix with integer entries (this follows from a construction in linear algebra known as the Cramer Rule). Thus, $B^{-1}(\mathbb{Z}^d) \subset \frac{1}{q}C(\mathbb{Z}^d) \subset \frac{1}{q}\mathbb{Z}^d$. So $\mathrm{Fix}(A) = B^{-1}(\mathbb{Z}^d)/\mathbb{Z}^d \subset \left(\frac{1}{q}\mathbb{Z}^d\right)/\mathbb{Z}^d \subset \mathbb{Q}^d/\mathbb{Z}^d$. □

Now, since $\mathrm{Per}(A) = \bigcup_{k\geq 1} \mathrm{Fix}(A^k)$, and $\det(A^k - \mathrm{id}) \neq 0$ for every k, we conclude that $\mathrm{Per}(A) \subset \mathbb{Q}^d/\mathbb{Z}^d$.

$\square$

8.4. Linear Algebra and Hyperbolicity

Let us investigate the condition that $\det(A - \mathrm{id}) \neq 0$. This implies that 1 is not an eigenvalue of A, so that there is no nonzero vector fixed by A. Now if $\det(A^n - \mathrm{id}) \neq 0$, then 1 is not an eigenvalue of A^n. Then there exists some λ, an eigenvalue of A, such that $\lambda^n = 1$. Any $\lambda \in \mathbb{C}$ satisfying $\lambda^n = 1$ for some $n \in \mathbb{N}$ is said to be a *root of unity*.

Definition 8.4.1. A matrix $A \in GL(d, \mathbb{Z})$ is said to be *hyperbolic* or *Anosov* if A does has no eigenvalue of modulus 1

Let us focus now on $\mathbb{T}^2$. We analyze the cases of A having a root of unity as an eigenvalue, and not having a root of unity as an eigenvalue separately. Note that every matrix A has two eigenvalues λ and μ, counting multiplicity. In this case, $\det(A) = \lambda\mu$, so $\mu = \pm\lambda^{-1}$.

If there is an eigenvalue λ which is a root of unity, we have the other eigenvalue (counting multiplicity) is $\pm\lambda^{-1}$. Now since $\lambda^n = 1$, $|\lambda| = 1$, so $\left|\lambda^{-1}\right| = 1$.

If $\det(A) = -1$, and λ is a root of unity, then $-\lambda^{-1} = -\overline{\lambda}$ is also an eigenvalue. But since the characteristic polynomial of a real matrix has real coefficients, its roots appear in complex conjugate pairs, so $\overline{\lambda}$ and $-\lambda$ must also be eigenvalues. Since there can be only two eigenvalues and $\lambda \neq -\lambda$, we conclude that $\lambda = \overline{\lambda}$, and λ is a real number of absolute value 1. Hence the eigenvalues must be ± 1.

In the case when $\det(A) = 1$, we have two eigenvalues λ and $\lambda^{-1} = \overline{\lambda}$. When A is a 2×2 matrix, its characteristic polynomial takes a particularly simple form. To define it, we need another notion from linear algebra:

Definition 8.4.2. If $A \in GL(d, \mathbb{R})$, let $\mathrm{Tr}(A) = \sum_{i=1}^{d} a_{ii}$, the sum of the diagonal entries. $\mathrm{Tr}(A)$ is called the *trace* of the matrix A.

The trace is an important number associated to a matrix, and has a particularly important relationship to the eigenvalus of the matrix, as we shall see.

Lemma 8.4.3. *If p is a monic polynomial of degree d with complex coefficients (ie, $p(x) = x^d + a_{d-1}x^{d-1} + \cdots + a_1 x + a_0$) and $\lambda_1, \ldots, \lambda_d$ (where we count roots with multiplicity, then:*

(1) $\prod_{i=1}^{d} \lambda_i = (-1)^d a_0$
(2) $\sum_{i=1}^{d} \lambda_i = -a_{d-1}$

Proof. Note that the Fundamental Theorem of Algebra gives us d roots (counting multiplicities). Thus, we can write:

$$\begin{aligned} p(x) &= \prod_{i=1}^{d} (x - \lambda_i) \\ &= x^d - \left(\sum_{i=1}^{d} \lambda_i\right) x^{d-1} + \cdots + (-1)^d \prod_{i=1}^{d} \lambda_i \end{aligned}$$

$\square$

Lemma 8.4.4. *If A is a $d \times d$ matrix and b_i are the coefficients of its characteristic polynomial, then $b_{d-1} = -\mathrm{Tr}(A)$ and $b_0 = (-1)^d \det(A)$.*

Proof. The case of the determinant is easy, since we evaluate the characteristic polynomial $p_A(\lambda) = \det(\lambda\,\mathrm{id} - A)$ at 0, we get the determinant of $-A$, or $(-1)^d \det(A)$. But the coefficient b_0 is exactly the constant coefficient (ie, the evaluation at 0).

We will prove the trace formula by induction. In the case of $d = 1$, this is trivial, since b_{d-1} is the only entry of the matrix, which is the trace.

Let us verify the inductive step. Recall that the trace and characteristic polynomial can be computed in block form by writing:

$$\det(\lambda \operatorname{id} - A) = \det \begin{pmatrix} \lambda - a_{11} & a_r \\ a_c & \lambda \operatorname{id} - \widehat{A} \end{pmatrix} = (\lambda - a_{11}) \cdot \det\left(\lambda \operatorname{id} - \widehat{A}\right) - a_r a_c$$

where $\widehat{A}$ is the $(d-1) \times (d-1)$ minor of A in the bottom right, and a_r and a_c are $(d-1)$-dimensional rows and columns, respectively. Then if $\hat{b}_i$ represent the coeffients of the characteristic polynomial of $\det(\lambda \operatorname{id} - \widehat{A})$, by induction, we have that $\hat{b}_{d-2} = -\operatorname{Tr}(\widehat{A}) = -\sum_{k=2}^{d} a_{kk}$.

Expressed this way, we have exactly two contributions to the λ^{n-1} term of p_A. First, if we distribute λ into $\det(\lambda \operatorname{id} - \widehat{A})$, we need to take the λ^{n-2} term, so we get $\hat{b}_{d-2}$. If we distribute $-a_{11}$, we need to take the λ^{n-1} term. But the polynomial is always monic, so we get $-a_{11}$. Thus,

$$b_{n-1} = \hat{b}_{d-2} - a_{11} = -a_{11} - \sum_{k=2}^{d} a_{kk} = -\sum_{k=1}^{d} a_{kk} = -\operatorname{Tr}(A)$$

$\square$

We get the following directly from the preceding lemmas:

Corollary 8.4.5. $\operatorname{Tr}(A)$ *is the sum of the eigenvalues and* $\det(A)$ *is the product of the eigenvalues*

Now, let $A \in SL(2, \mathbb{Z})$ so that again, $p_A(x) = x^2 - \operatorname{Tr}(A)x + \det(A) = x^2 - (\lambda + \overline{\lambda})x + 1$ be its characteristic polynomial. Since A is an integer matrix, and $|\lambda| = 1$, we know that $-2 \le \operatorname{Tr}(A) = \lambda + \overline{\lambda} \le 2$, so $\operatorname{Tr}(A) = 0, \pm 1, \pm 2$. For each trace, there are unique eigenvalues for which we can obtain it. This follows since if $\lambda, \overline{\lambda}$ are the eigenvalues of A, then $\lambda + \overline{\lambda} = k$ has exactly two solutions. Indeed, we can write $\lambda = x + iy$ as a complex number. Then $\lambda + \overline{\lambda} = 2x$, so x is uniquely determined. Once we know x there are exactly 2 values of y (if they exist) to solve the equation $x^2 + y^2 = 1$. We can thus fill in the following table of finite possibilities for the eigenvalues:

$\operatorname{Tr}(A)$	$\{\lambda, \lambda^{-1}\}$	Example Matrix
-2	$\{-1, -1\}$	$\begin{pmatrix} -1 & k \\ 0 & -1 \end{pmatrix}$
-1	$\left\{-\frac{1}{2} + i\frac{\sqrt{3}}{2}, -\frac{1}{2} - i\frac{\sqrt{3}}{2}\right\}$	$\begin{pmatrix} 1 & 1 \\ -1 & 0 \end{pmatrix}$
0	$\{i, -i\}$	$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$
1	$\left\{\frac{1}{2} + i\frac{\sqrt{3}}{2}, \frac{1}{2} - i\frac{\sqrt{3}}{2}\right\}$	$\begin{pmatrix} 0 & -1 \\ 1 & -1 \end{pmatrix}$
2	$\{1, 1\}$	$\begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix}$

We have thus shown that:

Proposition 8.4.6. *$A \in GL(2, \mathbb{Z})$ has an eigenvalue λ with $|\lambda| = 1$ if and only if either* $\det(A) = -1$ *and* $\operatorname{Tr}(A) = 0$ *or* $\det(A) = 1$ *and* $\operatorname{Tr}(A) \in \{-2, -1, 0, 1, 2\}$. *That is, A is hyperbolic if and only if* $|\operatorname{Tr}(A)| \ge 3$ *and* $\det(A) = 1$ *or* $|\operatorname{Tr}(A)| \ge 1$ *and* $\det(A) = -1$.

Let us look very closely at the dynamics of $A : \mathbb{T}^2 \to \mathbb{T}^2$ with $A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$. We know that $|\operatorname{Fix}(A^n)| = |\det(A^n - \operatorname{id})|$. But if A is a 2×2 matrix, its characteristic polynomial is $p_A(x) = \det(A - x \operatorname{id}) = x^2 - \operatorname{Tr}(A)x + \det(A)$. In our case, $x^2 - 3x + 1$.

So, if we wish to compute the number of fixed points we are interested in $|\mathrm{Fix}(A)| = |\det(A - \mathrm{id})| = |p_A(1)| = |2 - \mathrm{Tr}(A^n)|$.

Thus, in the case of the matrix A, we know that $\lambda + \lambda^{-1} = 3$, so $\mathrm{Tr}(A^n) = \lambda^n + \lambda^{-n}$ (this follows since the eigenvalues of the power is the power of the eigenvalues). Thus, $|\mathrm{Fix}(A^n)| = \lambda^n + \lambda^{-n} - 2$, growing exponentially fast.

8.4.1. Primary Decomposition and Higher Dimensions. To analyze the case of higher dimensions, we need another tool, called the *primary decomposition*. To describe such a decomposition, we need to expand our thinking about eigenspaces:

PROPOSITION 8.4.7. *λ is an eigenvalue if and only if $\ker(A - \lambda\,\mathrm{id}) \neq \{0\}$, and in this case $E_\lambda = \ker(A - \lambda\,\mathrm{id})$*

This proposition is basic linear algebra, and will not be proven. It gives us motivation to define:

DEFINITION 8.4.8. If λ is an eigenvalue of a linear transformation A, then the *generalized eigenspace* of λ is the space $E^\lambda = \bigcup_{n \geq 0} \ker(A - \lambda\,\mathrm{id})^n$. The *stable, unstable, and central* subspaces are the spaces:

$$\begin{aligned} E^s &= \bigoplus_{|\lambda|<1} E^\lambda \\ E^c &= \bigoplus_{|\lambda|=1} E^\lambda \\ E^u &= \bigoplus_{|\lambda|>1} E^\lambda \end{aligned}$$

THEOREM 8.4.9 (Primary Decomposition). *Each linear isomorphism $A : \mathbb{R}^d \to \mathbb{R}^d$ has a primary decomposition:*

$$\mathbb{R}^d = \bigoplus_{i=1}^{r} E^{\lambda_r} = E^s \oplus E^c \oplus E^u$$

SKETCH OF PROOF. The proof requires some tools from algebra, and we sketch it here. The *minimal polynomial* over $\mathbb{R}$ is the smallest-degree, monic polynomial p with coefficients in $\mathbb{R}$ such that $p(A) = 0$. For instance, if $A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$, then its minimal polynomial is $p(x) = x^2 - 3x + 1$, since:

$$p(A) = A^2 - 3A + 1 = \begin{pmatrix} 5 & 3 \\ 3 & 2 \end{pmatrix} - 3\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = 0$$

And any linear polynomial will not be able to kill the off-diagonal terms.

For notational convenience, let $\mathbb{R}[x]$ denote the space of all polynomials with coefficients in $\mathbb{R}$. Let $p(x) = p_1(x)^{j_1} p_2(x)^{j_2} \dots p_r(x)^{j_r}$ be the factorization of p into irreducible terms. That is, each p_i is a monic polynomial over $\mathbb{R}$, of the form $x - \lambda_i$ with $\lambda_i \in \mathbb{R}$ or $x^2 + \lambda_i$ with $\lambda_i > 0$. Take $q_i(x) = \prod_{k \neq i} p_k(x)^{j_k}$, so that $p(x) = p_i(x) q_i(x)$. We claim that if I is the set of polynomials of the form $f(x) = \sum_{i=1}^r f_i(x) q_i(x)$ with f_i being a polynomial with coefficients in $\mathbb{R}$, then I contains 1. It is easy to see that I is closed under multiplication and addition, so I is an *ideal*. So we use Lemma 8.4.10 to conclude that $I = \{g(x)f(x) : g(x) \in \mathbb{R}[x]\}$. We claim that $f(x)$ (called the *generator of* I) is 1. Indeed, $f(x)$ must divide each q_i, since $q_i \in I$, so $f(x) = p_i^{l_i}$ for some $l_i < j_i$. But $f(x)$ cannot be divisible by any p_i, since it must also divide $\prod_{k \neq i} p_k^{j_k}$. Hence $f = 1$ and $1 \in I$.

So we can write $1 = \sum_{i=1}^r g_i q_i$. Let $P_i = g_i(A) q_i(A)$, so that $\mathrm{id} = \sum_{i=1}^r P_i$. Furthermore, note that if $i \neq j$,

$$P_i \circ P_j = g_i(A)q_i(A)g_j(A)q_j(A) = g_i(A)g_j(A) \prod_{k \neq i} p_k^{j_k} \prod_{k \neq j} p_k^{j_k} = g_i(A)g_j(A)p(A) \prod_{k \neq i,j} p_k^{j_k} = 0.$$

So by Lemma 8.4.11, each P_i has eigenvalues only 1 or 0, and $\mathbb{R}^d = \bigoplus_{i=1}^r P_i(\mathbb{R}^d)$. Since P_i is a polynomial in A, we know that $P_i A = AP_i$ and if $E_i = P_i(\mathbb{R}^d)$, $AE_i = AP_i\mathbb{R}^d = P_i A\mathbb{R}^d = P_i\mathbb{R}^d = E_i$. Thus each E_i is invariant. We claim that E_i is the generalized eigenspace for the roots of p_i. Indeed, if $(A - \lambda\,\mathrm{id})^k v = 0$, then $(x - \lambda)^k$ divides p. Since p is the minimal polynomial, by Lemma 8.4.12, we get that any eigenvalue of $A|_{E_i}$ must be λ_i. □

We used three Lemmas in our proof, and will not provide proofs of them here. They are standard knowledge in linear algebra.

Lemma 8.4.10. *If I is an ideal in $\mathbb{R}[x]$, then $I = \{g(x)f(x) : g(x) \in \mathbb{R}[x]\}$ for some $f(x) \in \mathbb{R}[x]$*

Lemma 8.4.11. *If $P_i : \mathbb{R}^d \to \mathbb{R}^d$ are linear transformations such that $\mathrm{id} = \sum_{i=1}^r P_i$ and $P_i \circ P_j = 0$, then $P_i^2 = P_i$, and $\mathbb{R}^d = \bigoplus_{i=1}^r P_i(\mathbb{R}^d)$.*

Lemma 8.4.12. *For any linear transformation $A : \mathbb{R}^d \to \mathbb{R}^d$, the minimal polynomial divides the characteristic polynomial. Furthermore, the roots of the polynomials coincide.*

8.5. Stables and Unstables

In understanding the dynamics for the nonperiodic points, it will again be useful to use lifts. There is a canonical choice of lifts for automorphisms, namely, the corresponding linear map on $\mathbb{R}^2$. Note that $\widetilde{A}^n$ is diagonalizable, so we can understand the orbits by looking at the eigenspaces. One eigenspace will be expanded, the distance to the origin is growing exponentially at exponential rate λ. The other eigenspace is contracted exponentially to 0 with exponential rate of contraction λ.

If a point $x \in \mathbb{R}^2$ does not lie in some eigenspace, we can use the linearity and write $x = x^u + x^s$, where x^s is the contribution of the λ eigenspace and x^s is the contribution of the λ^{-1} eigenspace, since:

$$Ax = \lambda x^u + \lambda^{-1} x^s$$

Then the contribution of x^u will grow exponentially at a rate of λ, and the contribution of x^s will exponentially decay at a rate of λ. In particular, $\lim_{n\to\infty} d(\widetilde{A}^n x, \widetilde{A}^n x^u) \to 0$.

To see how this looks on the torus, we need to investigate how the eigenspaces look when projected. One can verify directly that $e^u = (1, \lambda - 2) \in E^u$ (where E^u is the eigenspace for λ) and $e^s = (1 - \lambda, \lambda) \in E^s$. Since they are one-dimensional, know that $E^u = \mathbb{R}e^u$ and $E^s = \mathbb{R}e^s$.

Proposition 8.5.1. *If $p : \mathbb{R}^2 \to \mathbb{T}^2$ is the projection and A is a hyperbolic toral automorphism, $p(x + E^u)$ and $p(x + E^s)$ is dense in $\mathbb{T}^2$ for every $x \in \mathbb{T}^2$*

In fact, this proposition is proved using the following:

Theorem 8.5.2. *Let $G \subset \mathbb{R}^d$ be a closed subgroup. Then there exist $k, l \geq 0$ and an injective linear map $L : \mathbb{R}^{k+l} \to \mathbb{R}^d$ such that $L(\mathbb{R}^k \times \mathbb{Z}^l) = G$. That is $L|_{\mathbb{R}^k \times \mathbb{Z}^l}$ is an isomorphism.*

The fact that the isomorphism comes from an injective linear map is an additional extra restriction. Consider the example of $\mathbb{Z} + \sqrt{2}\mathbb{Z} = \{a + b\sqrt{2} : a, b \in \mathbb{Z}\} \subset \mathbb{R}$. Then there is a linear map $\phi : \mathbb{R}^2 \to \mathbb{R}$ sending $(t, s) \mapsto t + \sqrt{2}s$, and $\phi|_{\mathbb{Z}^2}$ is an (algebraic) isomorphism. But ϕ cannot be made an injective linear map.

Proof of Theorem 8.5.2. Note that the case of $d = 1$ follows from Theorem 2.5.4. We proceed by induction. Let $G \subset \mathbb{R}^d$, and assume first that there is some $v \in G$ such that $\mathbb{R}v = \{tv : t \in \mathbb{R}\} \subset G$. Then if $E = (\mathbb{R}v)^{\perp} = \{w \in \mathbb{R}^d : \langle w, v \rangle = 0\}$, by elementary linear algebra:

(1) $E \cong \mathbb{R}^{d-1}$

(2) $E \oplus \mathbb{R}v \cong \mathbb{R}^d$

Let $p : \mathbb{R}^d \to E$ be the orthogonal projection. That is, if $v = tv + w'$, with $w' \in (\mathbb{R}v)^\perp$, then $p(v) = w'$. Let $F : \mathbb{R}^{d-1} \to E$ be any isomorphism onto its image, and $G_E = G \cap E$. Then $\widehat{G}_E = F^{-1}(G_E)$ is a closed subgroup of $\mathbb{R}^{d-1}$. So there exists $L_0 : \mathbb{R}^{k_0} \times \mathbb{R}^{l_0} \to \mathbb{R}^{d-1}$ such that $L_0(\mathbb{R}^{k_0} \times \mathbb{R}^{l_0}) = \widehat{G}_E$. Define the map $L : \mathbb{R}^{k_0+1} \times \mathbb{R}^{l_0} \to \mathbb{R}^d$ by:

$$L(x,t,y) = F(L_0(x,y)) + tv$$

Now we claim that $p(G) = G_E$. We claim that L is injective. Indeed, if $F(L_0(x,y)) + tv = 0$, then $F(L_0(x,y)) = -tv \in \mathbb{R}v \oplus (\mathbb{R}v)^\perp = \{0\}$. So we need only to show that $L|_{\mathbb{R}^k \times \mathbb{Z}^l}$ is onto G, where $k = k_0 + 1$ and $l = l_0$. If $z \in G$, the $p(z) \in G \cap E$ and there exists a unique t such that $z = p(z) + tv$. But then by induction, there exist unique x and y such that $L_0(x,y) = F^{-1}(p(z))$. Then $L(x,t,y) = z$.

Our last case is to consider the case when there is no $v \in G$ such that $\mathbb{R}v \subset G$. Then there exists $v \in G$ such that $tv \notin G$ for $0 < t < 1$. Let $G_v = G + \mathbb{R}v$, so that G_v falls into the previous case, and there is a linear map $L_v : \mathbb{R}^k \times \mathbb{R} \times \mathbb{R}^l \to \mathbb{R}^d$ such that $L_v(\mathbb{R}^k \times \mathbb{R} \times \mathbb{Z}^l) = G_v$. In our construction, we know that $L_v(0,1,0) = v$. We claim that $L_v(\mathbb{R}^k \times \mathbb{Z} \times \mathbb{Z}^l) = G$. We know that $L_v(\mathbb{R}^k \times \mathbb{Z} \times \mathbb{Z}^l) \subset G$, since $L_v(x,n,y) = F(L_0(x,y)) + nv$, $F \circ L_0(\mathbb{R}^k) = G_E \subset G$ and $nv \in G$ since $v \in G$.

To see the other inclusion, suppose that $z \in G \subset G_v$. Then since $L_v : \mathbb{R}^k \times \mathbb{R} \times \mathbb{Z}^l)$ is an isomorphism onto G_v, there exists a unique (x,t,y) such that $L_v(x,t,y) = z$. Let $z' = z - F(L_0(x,y)) - \lfloor t \rfloor v = (t - \lfloor t \rfloor)v$. We know that $t - \lfloor t \rfloor \in [0,1)$, but by construction there exists no $s \in (0,1)$ such that $sv \in G$. Then $t - \lfloor t \rfloor = 0$, and $t \in \mathbb{Z}$. That is $L(\mathbb{R}^k \times \mathbb{Z} \times \mathbb{Z}^l) \supset G$. □

Proof of Proposition 8.5.1. Let $L \subset \mathbb{R}^2$ be any line, and $H = \overline{p(L)}$. Then $\widetilde{H} = p^{-1}(H) = \overline{L + \mathbb{Z}^2} \subset \mathbb{R}^2$ is a closed subgroup. It must contain both L and $\mathbb{Z}^2$, so it must be isomorphic one of $L \times \mathbb{Z}$ or $\mathbb{R}^2$ by Theorem 8.5.2. If it is isomorphic to $L \times \mathbb{Z}$, then L together with its translates in some direction contain $\mathbb{Z}^2$. We claim that this implies that there is a point of $\mathbb{Z}^2$ in L itself. Indeed, if L contained no integral point, then we can choose some $x_0 \in \mathbb{R}^2$ such that x_0 generates the copy of $\mathbb{Z}$ in the isomorphism, (ie, if $\varphi : \widetilde{H} \to L \times \mathbb{Z}$ is an isomorphism of Theorem 8.5.2, $x_0 = \varphi^{-1}(0,1)$). Then $(0,1) = tz + nx_0$ and $(1,0) = sz + mx_0$, where z is any point on the line L. Then observe that:

$$(tm - ns)z = (-n, m)$$

$tm - ns$ cannot be 0, since $(0,1)$ and $(1,0)$ do not lie on the same line (ie, the transformation $\begin{pmatrix} t & s \\ n & m \end{pmatrix}$ is nondegenerate, since x_0 is not on the line L). But then $(-n,m)$ is a point of $\mathbb{Z}^2$ on the line L, since $(tm - ns)z \in L$. So L contains no integral point.

We have two options for finishing the proof, as all that is left to show is that the slope of the eigenspaces is irrational. This can be observed directly from computation. We can also proceed by contradiction. First, note that if $(a,b) \in \mathbb{Z}^d \cap E^u$, then applying A^{-1} many times, this point will tend to 0. But it must always be an integer, so it must be 0 to begin with. Thus, the slope must be irrational.

Another way to prove this would be conclude that if $(a,b) \neq (0,0)$ was in L, the curve $\gamma(t) = (ta, tb)$ would lie in L completely. Projecting to the torus, this would correspond to a closed curve. But taking a sufficiently small backward iterate, this curve will eventually not be an integer (since it is contracted by the dynamics A), so it will not be closed. Since closed curves are taken to closed curves under a continuous map, we again arrive at a contradiction. □

The proof of Proposition 8.5.1 carries over nicely as stated, but can be extended to an event greater generality!

DEFINITION 8.5.3. Let $A : \mathbb{T}^d \to \mathbb{T}^d$ be a toral automorphism, and $\widetilde{A} : \mathbb{R}^d \to \mathbb{R}^d$ be the linear lift. A is said to be *ergodic* if $\widetilde{A}$ has no eigenvalues which are roots of unity.

EXERCISE 8.5.1. Observe that every hyperbolic matrix is ergodic. Show that the opposite is true in dimensions 2, 3 and 5. Find a counterexample in dimension 4.

THEOREM 8.5.4. *Let $A_0 : \mathbb{T}^d \to \mathbb{T}^d$ be a toral automorphism with real eigenvalues $\lambda_1, \dots, \lambda_r$ with associated linear transformation $A : \mathbb{R}^d \to \mathbb{R}^d$. Then:*

(1) $\mathbb{R}^d = E^s \oplus E^c \oplus E^u$
(2)

$$\begin{aligned} E^s &= \left\{v \in \mathbb{R}^d : |A^n v| \to 0 \text{ as } n \to \infty\right\} \\ E^c &= \left\{v \in \mathbb{R}^d : |A^n v|^{1/|n|} \to 1 \text{ as } n \to \pm\infty\right\} \\ E^u &= \left\{v \in \mathbb{R}^d : \left|A^{-n} v\right| \to 0 \text{ as } n \to \infty\right\} \end{aligned}$$

(3) If A_0 is ergodic, then $p(E^*)$ is dense in $\mathbb{T}^d$, for $* = s, u$
(4) If A_0 is ergodic, then A_0 is topologically transitive

PROOF. The first claim is a restatement of Theorem 8.4.9. We compute the second in the case when $\mathbb{R}^d$ has a basis of eigenvectors. It follows more generally by observing that if $v \in E^\lambda$ and $(A - \lambda \,\mathrm{id})^r|_{E^\lambda} = 0$, then A^r can be written in terms of a polynomial of lower degree in A. For instace, if $A = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}$, $(A - \lambda \,\mathrm{id})^2 = 0$, so $A^2 = 2\lambda A - \lambda^2 \,\mathrm{id}$. In this case, we can write:

$$A^n e_2 = \lambda^n e_2 + 2 \sum_{k=0}^{n-1} \lambda^k e_1$$

A similar statement can be made in the general case (see Lemma 8.6.4). We will still get the exponential contributions in our subsequent computations, and any other contributions from lower powers will decay.

So we will continue with the assumption that every $v \in E^\lambda$ is an eigenvector with eigenvalue λ. Note that the inclusions $E^* \subset \left\{v \in \mathbb{R}^d : |A^n v| \to 0 \text{ as } n \to \pm\infty\right\}$ (and a similar statement for the central subspace) follow trivially from the dynamics. To see the other inclusions, note that if v is a vector in the right hand side and Δ is the set of eigenvalues of A, we can write $v = \sum_{\lambda \in \Delta} v^\lambda$, with $v^\lambda \in E^\lambda$, by Theorem 8.4.9. Then if $n \in \mathbb{Z}$:

$$\begin{aligned} |A^n v| &= \left| A^n \left(\sum_{\lambda \in \Delta} v^\lambda \right) \right| \\ &\geq \sum_{\lambda \in \Delta} \left|A^n v^\lambda\right| \\ &= \sum_{\lambda \in \Delta} |\lambda|^n \left|v^\lambda\right| \end{aligned}$$

If this tends to 0 as $n \to \infty$, then $v^\lambda = 0$ whenever $\lambda \geq 1$. That is, $v \in E^s$. Similary, if it tends to 0 as $n \to -\infty$, $v^\lambda = 0$ whenever $\lambda \leq 1$. That is, $v \in E^u$.

Now, for the case of E^c, note that if for any $\lambda_0 \in \Delta$ for which λ_0 and $n \geq 0$:

$$|A^n v|^{1/n} \geq \left(\sum_{\lambda \in \Delta} |\lambda^n| \left|v^\lambda\right| \right)^{1/n} \geq \left(|\lambda_0|^n \left|v^{\lambda_0}\right|\right)^{1/n} \to |\lambda_0|$$

Similarly, if we use $n \leq 0$, we get $|A^n v|^{-1/n} \geq \left(|\lambda_0|^n \left|v^{\lambda_0}\right|\right)^{-1/n} \to |\lambda_0|^{-1}$. Thus, if the limit must be 1, we know that $|\lambda_0| = 1$ for every λ with $v^\lambda \neq 0$, and $v \in E^c$.

We need to use Theorem 8.5.2 for the next statement. Note that $p(E^*)$ is dense in $\mathbb{T}^d$ if and only if $\overline{\mathbb{Z}^d + E^*} = \mathbb{R}^d$. Since E^* is a subspace, $\mathbb{Z}^d + E^*$ is a subgroup. Thus, $\overline{\mathbb{Z}^d + E^*}$ is a closed subgroup, since the closure of any subgroup is again a subgroup. Then there exists some injective linear map $L : \mathbb{R}^{k+l} \to \mathbb{R}^d$ such that $L|_{\mathbb{R}^k \times \mathbb{Z}^l}$ is an isomorphism onto G by Theorem 8.5.2. Since $G \supset \mathbb{Z}^d$, $k + l = d$. It follows since $k + l \leq d$ by injectivity, and the image of L must contain $\mathbb{Z}^d$, and hence a basis.

Thus, $L : \mathbb{R}^{k+l} \to \mathbb{R}^d$ is an isomorphism. If we can thus show that $l = 0$, we are done. Note that $A' = L^{-1} \circ A \circ L$ is another linear map. Furthermore, since $G' = L^{-1}(G) = \mathbb{R}^k \times \mathbb{Z}^l$, and G is invariant under A, G' is invariant by A'.

Let us first deal with $* = s$, the E^u case also follows by similar argument. Since E^s is connected, it must be contained in the connected component of 0, or $\mathbb{R}^k \times \{0\}$. Since $A'(0) = 0$, by connectedness, again we get $A'(\mathbb{R}^k \times \{0\}) \subset \mathbb{R}^k \times \{0\}$. Finally, since A' is a linear isomorphism, we get $A'(\mathbb{R}^k \times \{0\}) = \mathbb{R}^k \times \{0\}$.

Thus, A' induces a linear map $B : \mathbb{R}^l \to \mathbb{R}^l$, since $\mathbb{R}^k \times \{0\}$ is invariant and we identify $\mathbb{R}^l \cong \mathbb{R}^{k+l}/(\mathbb{R}^k \times \{0\})$ (this construction follows from abelian group qoutients and induced homomorphisms). Since A' is an isomorphism, so is B (note that this does not hold for groups in general, but only because we quotient by a vector subspace).

We claim that $\mathbb{Z}^l \subset \mathbb{R}^l$ is invariant by B. This follows because A' is an isomorphism $\mathbb{R}^k \times \mathbb{Z}^l \to \mathbb{R}^k \times \mathbb{Z}^l$, and the linear map B is actually the extension of the map induced by $A' : (\mathbb{R}^k \times \mathbb{Z}^l)/\mathbb{R}^k \to (\mathbb{R}^k \times \mathbb{Z}^l)/\mathbb{R}^k \cong \mathbb{Z}^l$.

Since $\mathbb{R}^l$ naturally sits inside $\mathbb{R}^{k+l}$, any eigenvalue of B will also be an eigenvalue of A'. Since $L^{-1}(E^s) \subset \mathbb{R}^k \times \{0\}$, all generalized eigenspaces with eigenvalues less than 1 are contained in $\mathbb{R}^k \times \{0\}$. Since B also induces an automorphism of $\mathbb{T}^l$, $|\det(B)| \geq 1$. Since the product of the eigenvalues must have modulus 1, and each eigenvalue must have modulus greater than 1, we conclude that each has modulus exactly 1.

By a problem in the homework, this implies that each of these eigenvalues are roots of unity, a contradiction of ergodicity. $\square$

8.6. Topological Mixing

Ergodic linear automorphisms share one more imporant property:

Definition 8.6.1. A dynamical system $f : X \to X$ is *topologically mixing* if for every pair of open sets U, V, there exists some N such that if $n \geq N$, $f^n(U) \cap V \neq \emptyset$

From Theorem 3.3.2, every topologically mixing system is topologically transitive. Is the opposite true in general?

Theorem 8.6.2. *Let $A : \mathbb{T}^d \to \mathbb{T}^d$ be an automorphism. The following are equivalent:*

(1) A is ergodic
(2) $p(E^s)$ and $p(E^u)$ are dense in $\mathbb{T}^d$
(3) A is topologically transitive
(4) A is topologically mixing

(1) $\implies$ (2) follows from our arguments before, so let us turn to (2) $\implies$ (3), (4). We will use the equivalent definition of transitivity using open sets provided by Theorem 3.3.2. Suppose that $p(E^s)$ and $p(E^u)$ are dense in $\mathbb{T}^d$, and let $U, V \subset \mathbb{T}^d$ be open sets. We will use the following Lemmas:

Lemma 8.6.3. *For every $\varepsilon > 0$, there exists some $R > 0$ such that if $x \in \mathbb{R}^d$ and $E^u_r = E^u \cap B_R(0)$, then $p(x + E^u_r) \cap B_v e(y) \neq \emptyset$ for every $y \in \mathbb{T}^d$ and $r \geq R$*

Lemma 8.6.4. *If $V \subset \mathbb{R}^d$ is a vector subspace, and $A : V \to V$ is a linear transformation whose eigenvalues are all greater than 1, then there exists $\lambda > 1$ and $N > 0$ such that if $n \geq N$, $A^n(B_r(0)) \supset B_{\lambda^n r}(0)$*

Note that taking the n^{th} power is necessary, consider the example $A = \begin{pmatrix} 5/4 & -5/4 \\ 0 & 5/4 \end{pmatrix}$. Then the only eigenvalue of A is $5/4$ and the point $v = (1,1)$ is sent to $(0, 5/4)$. Then $||Av|| = 5/4 < \sqrt{2} = ||v||$. We will prove the Lemmas after concluding a proof of the following proposition, which takes care of $(2) \implies (3), (4)$

Proposition 8.6.5. *If $p(E^u)$ is dense in $\mathbb{T}^d$, then f is topologically mixing*

Proof. Let U and V be open sets. Choose $\varepsilon > 0$ and $y \in V$ such that $B_\varepsilon(y) \subset V$. Let R be as in the lemma, and take any point $z \in U$ and choose $\delta > 0$ such that $z + p(E^u_\delta) \subset U$. Now, $A^n(z + p(E^u_\delta)) = A^n z + A^n(p(E^u_\delta)) = A^n z + p(A^n E^u_\delta)$. But if λ is the smallest eigenvalue of A which is greater than 1, by Lemma 8.6.4, $A^n E^u_\delta \supset E^u_{\lambda^n \delta}$.

This implies that $A^n U \supset A^n z + E^u_{\lambda^n \delta}$. By Lemma 8.6.3, such a set must intersect any open set V for suitably large n. That is, A is topologically mixing. □

Proof of Lemma 8.6.3. We proceed by contradiction. That is, we assume that there is some $\varepsilon_0 > 0$ such that for every $R > 0$, there is some $r \geq R$ and x_r, y_r such that $(x_r + p(E^u_r)) \cap B_{\varepsilon_0}(y_r) = \emptyset$. This implies that $p(E^u_r) \cap B_{\varepsilon_0}(y_r - x_r) \neq \emptyset$, so let $z_r = y_r - z_r$. Let r take values in $\mathbb{N}$, so that we can take a convergent subsequence $z_{r_i} \to z \in \mathbb{T}^2$. We claim that this implies $d(z, p(E^u)) \geq \varepsilon_0$ (contradicting the density of $p(E^u)$). Let $w \in p(E^u)$. Then there exists an n such that $i \geq n$ implies $w \in E^u_{r_i}$. Since $B_{\varepsilon_0}(z_{r_i}) \cap p(E^u_r) = \emptyset$, $d(z, w) = \lim_{i\to\infty} d(z_{r_i}, w) \geq \varepsilon_0$. We arrive at our contradiction and conclude the Lemma. □

Proof of Lemma 8.6.4. We will use the Primary Decomposition. It tells us that if $A : V \to V$ has eigenvalues which all have modulus greater than 1, then $V = \bigoplus_{i=1}^k E^{\lambda_i}$. Furthermore, there is an $n_i \geq 0$ such that $(A - \lambda_i)^{n_i}|_{E^{\lambda_i}} = 0$. Let $N_i = (A - \lambda_i \,\mathrm{id})|_{E^{\lambda_i}} : E^{\lambda_i} \to \mathbb{R}^d$, and $N : \mathbb{R}^d \to \mathbb{R}^d$ be the map $N(x_1, \ldots, x_k) = (N_1 x_1, \ldots, N_k x_k)$. Since the primary decomposition is a direct sum, this is a well-defined homomorphism. Furethermore, if $S_i = (A - N_i)|_{E^{\lambda_i}} = (\lambda_i \,\mathrm{id})|_{E^{\lambda_i}}$, and $S : \mathbb{R}^d \to \mathbb{R}^d$ is defined by $S(x_1, \ldots, x_k) = (Sx_1, \ldots, Sx_k)$, then $A = S + N$ (S is often called the *semisimple* part of A and N is the *nilpoent part* of A). By construction, $N^k = 0$ for some sufficiently large k, and S has a basis of eigenvectors (since it acts by multiplication in each primary summand). Furthermore, $NS = SN$, since both are constructed as polynomials in A and id.

Define the *conorm* of a linear transformation as:

$$m(A) = \min\left\{|Av| : v \in \mathbb{R}^d, |v| = 1\right\}$$

We can restate the Lemma as $m(A^n) > \lambda^n$ for suffiently large n. Observe that since $NS = SN$,

$$A^n v = (S + N)^n v = \sum_{i=0}^{n} \binom{n}{i} S^{n-i} N^i v = \sum_{i=0}^{k-1} \binom{n}{i} S^{n-i} N^i v = S^{n-k} \sum_{i=0}^{k-1} \binom{n}{i} S^{k-i} N^i v$$

Now observe that since S is invertible, $A'_n = \sum_{i=0}^{k-1} \binom{n}{i} S^{k-i} N^i$ is also an invertible linear transformation. Thus, there is a $\varepsilon_n > 0$ such that $B_{\varepsilon_n}(0) \subset A'_n B_1(0)$. We claim that there is a ε that works for every n. To see this, suppose otherwise. Then $\bigcap_{n\geq 0} A'_n B_1(0)$ does not contain a neighborhood of the identity. There then exist $v_n \in B_1(0)$ such that $A'_n v_n \to 0$. This is clearly impossible since the A'_n are integral combinations of finitely many fixed linear transformations $S^{k-i} N^i$ for $i = 0, \ldots, k-1$, and these will tend to 0 if and only if they eventually become 0. This would imply that $v_n = 0$ for sufficiently large n. So we conclude that $B_\varepsilon(0) \subset A'_n B_1(0)$ for every $n \in \mathbb{N}$.

Let λ be the square root of the smallest modulus of the eigenvalue of A so that $\lambda > 1$. Then choose $N \geq 2k$ large enough so that $\lambda^N \varepsilon > 1$. Then if $n \geq N$ and $|v| = 1$:

$$|A^n v| = \left|S^{n-k} A'_n v\right| \geq \lambda^n \left|A'_n v\right| \geq \lambda^n$$

□

It is obvious that topological mixing implies topological transitivity (that is (4) $\implies$ (3)), so the last step is to show that topologically transitivity implies ergodicity ((3) $\implies$ (1)). We proceed again by contradiction. Assume that A is not ergodic. Then there is an eigenvalue which is a root of unity. Note that A^T (the transpose of the matrix A) also has a root of unity as an eigenvalue. So there is some $v \in \mathbb{Z}^d$ and $k \in \mathbb{N}$ such that $(A^T)^k v = v$. We can assume that $v \in \mathbb{Z}^d$ since the $(A^T)^k$ has an eigenvalue of modulus 1 by construction, and if $((A^T)^k - \mathrm{id})v = 0$, the Gaussian elimination method will produce a solution with rational entries. Clearing the denominator gives an entry in $\mathbb{Z}^d$. Let $\varphi : \mathbb{T}^d \to \mathbb{R}$ be the function:

$$\varphi(x) = \sin\left(2\pi \langle x, v\rangle\right)$$

since this function is $\mathbb{Z}^d$-periodic on $\mathbb{R}$, it is well-defined on $\mathbb{T}^d$. Note also that $\varphi(A^k x) = \sin(2\pi \left\langle A^k x, v\right\rangle) = \sin(2\pi \left\langle x, (A^T)^k v\right\rangle) = \sin(2\pi \langle x, v\rangle) = \varphi(x)$. Now define $\eta(x) = \sum_{i=0}^{k-1} \varphi(A^i x)$. Then $\eta(Ax) = \eta(x)$ by construction. Choose $a \neq b \in \eta(\mathbb{T}^d)$ (Why is this possible?) and $\varepsilon > 0$ such that $B_\varepsilon(a) \cap B_\varepsilon(b) = \emptyset$. Then $U = \eta^{-1}(B_\varepsilon(a))$ and $V = \eta^{-1}(B_\varepsilon(b))$ are both invariant under A, since if $x \in U$, $\eta(x) = a$. But then $\eta(Ax) = a$ and $Ax \in U$. But then $A^n U \cap V = U \cap V = \emptyset$, so A is not topologically transitive.

CHAPTER 9

Entropy

We have seen many ways of trying to establish when two dynamical systems are the same, including rotation number degree for expanding maps, and the determinant of homotopy type. In this last chapter, we will develop a technique that will allow us to *distinguish* dynamical systems [1].

The technique we will use is orbit counting. In any dynamical system on a compact, infinite space, there will be uncountably many orbits (Why?), so we first need to develop a new way to count orbits:

Definition 9.0.6. Let X be compact and $T : X \to X$ a dynamical system. $x, y \in X$ are said to be (ε, n) *dependent* if

$$d(T^k x, T^k y) < \varepsilon \text{ for every } k = 0, 1, \dots, n-1$$

A set $C \subset X$ is said to be (ε, n)-independent if for every distinct pair $x, y \in C$, x and y are not (ε, n)-dependent.

This notion can be thought of as distinguishing orbits with a given precision: two points x and y are considered "the same" under the dynamical system T with precision ε up to time n if they are (ε, n)-dependent.

There are some natural questions:

(1) Are there equivalent definitions for (ε, n)-dependence?
(2) Are there equivalent definitions for (ε, n)-independence?
(3) Can you have an infinite (ε, n)-independent set? (That is, can we have infinitely many types of orbits up to finite precision?)
(4) How should we expect the size of (ε, n)-independent sets to vary with the parameters ε and n? (That is, how many orbits should we expect as we vary the length of the orbit and our precision?)

The answers to these questions will lead us to the notion of topological entropy.

9.1. Counting metrics

Let (X, d) be a metric space and $T : X \to X$ be a dynamical system. Define the following metrics:

$$d_T^n(x, y) = \max_{k=0,\dots,n-1} \left\{ d(T^k x, T^k y) \right\}$$

We get the following Proposition:

Proposition 9.1.1.

(a) d_T^n *is a metric for every* $n \in \mathbb{N}$
(b) x *and* y *are* (n, ε)*-dependent if and only if* $d_T^n(x, y) < \varepsilon$
(c) $B_\varepsilon^{d_T^n}(x) = \bigcap_{k=0}^{n-1} T^{-k}(B_\varepsilon(T^k(x))$
(d) d_T^n *and* d *induce the same topology on* X
(e) $C \subset X$ *is* (n, ε)*-independent if and only if for every distinct pair* $x, y \in X$, $x \notin B_\varepsilon^{d_T^n}(y)$

[1] In fact, this tool in the context of measurable dynamics can be the unique invariant of dynamical systems when a positive real!

PROOF.

(a) That d_T^n is a metric will follow from properties of d. First, since $d \geq 0$, then taking the maximum value of d will also yield a result greater than 0. If $d_T^n(x,y) = 0$, then since we take the maximum, we get that $d(T^k x, T^k y) = 0$ for every k. In particular, for $k = 0$ and hence since d is a metric, $x = y$.

Symmetry obviously follows from the symmetry of d. To verify the triangle inequality, note that:

$$d_T^n(x,z) = \max_{k=0,\ldots,n-1} \left\{ d(T^k x, T^k z) \right\} \leq \max_{k=0,\ldots,n-1} \left\{ d(T^k x, T^k y) + d(T^k y, T^k z) \right\} \leq \max_{k=0,\ldots,n-1} \left\{ d(T^k x, T^k y) \right\} + \max_{k=0,\ldots,n-1} \{ d$$

Thus, d_T^n is a metric.

REMARK 1. Observe that we did not use the continuity of T. This will appear when we show that d_T^n induces the same topology as d

□

This Proposition answers a few of our questions. It also shows that to estimate the number of distinct (n, ε) orbits, we can equivalently count the number of disjoint d_T^n-balls we can fit inside X. Such a set is called separated:

DEFINITION 9.1.2. Let (X, d) be a metric space. A ε-separated set is a subset $C \subset X$ such that if $x, y \in C$, $x \notin B_\varepsilon(y)$.

The following Proposition tells us that there are only finitely many (ε, n)-orbits, negatively answering our third question:

PROPOSITION 9.1.3. *If (X, d) is a compact metric space and ε is fixed, then the number of points in a ε-separated set is bounded*

PROOF. Suppose that there existed subsets C_k such that $|C_k| = k$ and C_k was ε-separated (this can always be arranged if the sets become arbitrarily large, by choosing the first k elements of the first set with more than k elements). We will derive a contradiction.

We will first show that if the sets C_k exist, then there is an infinite ε-separated set. Write $C_k = \left\{ x_1^{(k)}, x_2^{(k)}, \ldots, x_k^{(k)} \right\}$. Since X is compact the sequence $x_1^{(k)}$ (indexed by k) has a convergent subsequence, $x_1^{(k_l)} \xrightarrow{l \to \infty} x_1$. Repeat this process for the sequence $x_2^{(k_l)}$ to get a subsequence $x_2^{(k_{l_m})} \to \xrightarrow{m \to \infty} x_2$. We can iterate this construction to get a sequence x_k.

We claim that the $\{x_k\}$ is ε-separated. To see this, fix x_m, x_n with $m < n$. Then the subsequence chosen to get x_n is a subsequence of the subsequence chosen for x_m. That is if $x_m^{(k_l)} \to x_m$, there is a subsequence $x_n^{(k_{l_i})} \to x_n$. But by construction, $d(x_m^{(k_{l_i})}, x_n^{(k_{l_i})}) \geq \varepsilon$, so $d(x_m, x_n) \geq \varepsilon$.

We have thus shown that if the ε-separated sets can become arbitrarily large, then there must be an infinite ε-separated set. We claim that the sequence x_n has no convergent subsequence (this will contradict compactness of X). Indeed, suppose that x was an accumulation point of x_n. Then there exists some $N \in \mathbb{N}$ such that if $k \geq N$, $d(x_{n_k}, x) < \varepsilon/2$. But then $d(x_{n_N}, x_{n_{N+1}}) < d(x_{n_N}, x) + d(x, x_{n_{N+1}}) < \varepsilon$, which cannot happen because $\{x_n\}$ is ε-separated. □

Since d_T^n induces the same topology as d, there exists a size of ε-separated sets with respect to the metric d_T^n, $C(T, n, \varepsilon)$.

EXAMPLE 9.1.1. Consider the case of $T = L_m$ with $m \geq 2$, and compute $C(L_m, n, 1/m)$. We claim that $C(L_m, n, 1/m) = m^n$. To see this, we note that $B_{1/m}^{d_{L_m}^n}(x) = (x - m^{-n}, x + m^{-n}) \bmod 1$. We verify this by induction. When $n = 1$, this is trivial. Suppose that $d_{L_m}^{n+1}(x,y) < 1/m$. Then, in particular, $d_{L_m}^n(x,y) < 1/m$, so by the inductive hypothesis, $d(x,y) < m^{-n}$. Since m^{-n} is within the expanding constant, we get that $m^n d(x,y) \leq m^{n-1} d(L_m x, L_m y) < 1/m$. And we thus conclude that $d(x,y) < m^{-(n+1)}$.

Now we verify that $C(L_m, n, 1/m) = m^n$. Indeed, observe that if we take the set $C = \{0, 1/m^n, \ldots, (m^n-1)/m^n\}$, then C is $1/m$-separated with respect to $d^n_{L_m}$, since $d^n_{L_m}(i/m^n, (i+1)/m^n) = m^{n-1} \cdot 1/m^n = 1/m$ (similar computations hold for nonconsecutive elements).

To verify that this is the maximal number, suppose that C is any subset of $\mathbb{R}/\mathbb{Z}$ which is $1/m$-separated for $d^n_{L_m}$. Choose representatives in $[0,1)$ and order the subset as $C = \{c_0, \ldots, c_{k-1}\}$, with $c_i < c_{i+1}$. Then $d^n_{L_m}(c_i, c_{i+1}) \geq 1/m$, so $c_{i+1} - c_i \geq m^{-n}$. But then $c_{k-1} \geq (k-1)/m^n$, and also $c_k \leq (m^n-1)/m^n$, since $d^n_{L_m}(c_k, 0) \geq 1/n$. It then follows that $k \leq m^n$.

Example 9.1.1 motivates the expectation that the number of (ε, n)-orbits should grow exponentially with n. The exponential constant with which this grows varies with m, and coincidentally (or perhaps, not so coincidentally), coincides with the degree of the map! How can we pick out this constant? Let:

$$h(T, \varepsilon) = \liminf_{n\to\infty} \frac{1}{n} \log C(T, n, \varepsilon)$$

Observe that in our example, $h(L_m, 1/m) = \lim_{n\to\infty} \frac{1}{n} \log m^n = \lim_{n\to\infty} \log m = \log m$, which is different for every m. To remove the ε dependence, we take another limit:

$$h_{\mathrm{top}}(T) = \lim_{\varepsilon\to 0} h(T, \varepsilon) = \lim_{\varepsilon\to 0} \liminf_{n\to\infty} \frac{1}{n} \log C(T, n, \varepsilon)$$

$h_{\mathrm{top}}(T)$ is called the *topological entropy* of T and is a powerful tool in the analysis of topological dynamical systems.