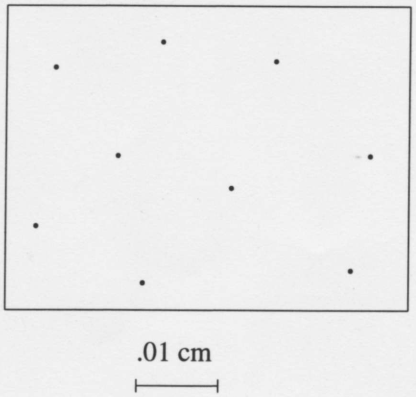
Entropy and Chaos in non-autonomous and autonomous systems

The previous two sections tell us that unless we impose strong restrictions on the functions f_i in a non-autonomous system, then we can not expect simple behavior of the orbits such as periodic orbits or fixed points. Consequently, in this section we explore methods of measuring how disordered the orbits of the dynamical system are. Specifically, the idea is that the rate at which nearby orbits spread away from each other is a measure of complexity.

One may wonder about the utility of this notion for neural computation. Topological entropy (a mathematical notion that measures the spreading out of orbits) measures how fast the dynamical system explores the whole space within an ϵ mesh. Since all the training algorithms are essentially search routines, entropy is one way to measure the quality of the search. Further, as described in section V, a neural net algorithm uses the exponential spreading of orbits to find a global minima.

The notion of orbits spreading out is introduced with a simple example. For a fixed distance of .01 centimeters and in a fixed area of space, the problem is to determine the maximum number of points that can be placed in this area so that no two points are closer than .01 centimeters apart. If we think of our whole space as the fixed rectangle, then 9 points can be placed so that all points are at a distance of at least .01 cm apart.

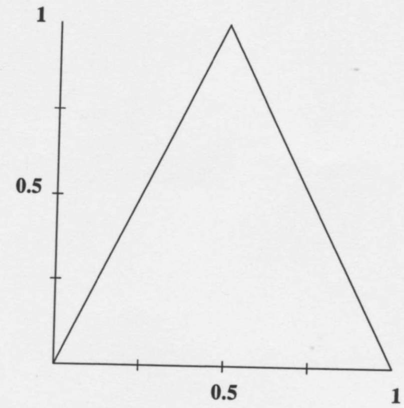


There is another way to think about points being placed in a region so they are not too close to each other. Suppose the resolution of a magnifying glass is 0.01 cm. If two points are less than 0.01 cm apart, then they cannot be distinguished. They appear to be one blob.

To extend this idea to a function f, we determine the number of points that can be packed in a space so that any two of them separate at least once by at least .01 cm. after no more than, say, 5 iterations of f. Notice that the emphasis is on separating by .01 cm at least once; e.g. it is acceptable if two points x, y are within .01 cm on the first 4 iterates as long as they separate the required amount by the fifth iterate.

A particular instance of two points separating by .01 cm. in at most 5 iterates is the following: The distance between x, y is less than .00001 cm. The distance between $f(x), f(y)$ is less than .00001 cm. The distance between $f \circ f(x), f \circ f(y)$ is less than .00001 cm. The distance between $f^3(x), f^3(y)$ is less than .00001 cm. The distance between $f^4(x), f^4(y)$ is less than .00001 cm. And the distance between $f^5(x), f^5(y)$ is .01 cm.

To illustrate, consider the tent map: $T : [0,1] \longrightarrow [0,1]$ and $T(x) = 2x$ if $x \in [0, 0.5]$ and $T(x) = 2 - 2x$ if $x \in (0.5, 1]$.



Fix the distance to be 0.5, and ask how many points can we pack into $[0,1]$ so that any two points separate by a distance of at least 0.5 in no more than 0 iterates. Clearly, the maximum number is at least three because the set $\{0.0, 0.5, 1.0\}$ satisfies the condition. On the other hand, any set with four points has two of them less than or equal to 0.5 units apart; this violates the separability condition.

The maximum number for 0, and 1 iterates is five points. This can be seen with the set $\{0, \frac{1}{4}, \frac{2}{4}, \frac{3}{4}, 1\}$. For example, the distance between $T(0)$ and $T(\frac{1}{4})$ is $\frac{1}{2}$, so they separate. Since there are so many points, the following diagram makes it easier to see that this set satisfies the condition:

$0 \xrightarrow{f} 0$

$\frac{1}{4} \xrightarrow{f} \frac{2}{4}$

$\frac{2}{4} \xrightarrow{f} 1$

$\frac{3}{4} \xrightarrow{f} \frac{2}{4}$

$1 \xrightarrow{f} 0.$

The answer for $0, 1$, or 2 iterates is given by the set $\{0, \frac{1}{8}, \frac{2}{8}, \frac{3}{8}, \frac{4}{8}, \frac{6}{8}, \frac{7}{8}, 1\}$; it has a maximum number of elements that separate by at least 0.5 units in $0, 1$, or 2 iterates. This assertion is clear as one sees by comparing any two points in the diagram:

$0 \xrightarrow{f} 0 \xrightarrow{f} 0$

$\frac{1}{8} \xrightarrow{f} \frac{2}{8} \xrightarrow{f} \frac{4}{8}$

$\frac{2}{8} \xrightarrow{f} \frac{4}{8} \xrightarrow{f} \frac{8}{8}$

$\frac{3}{8} \xrightarrow{f} \frac{6}{8} \xrightarrow{f} \frac{4}{8}$

$\frac{4}{8} \xrightarrow{f} \frac{8}{8} \xrightarrow{f} 0$

$\frac{6}{8} \xrightarrow{f} \frac{4}{8} \xrightarrow{f} 1$

$\frac{7}{8} \xrightarrow{f} \frac{2}{8} \xrightarrow{f} \frac{4}{8}$

$1 \xrightarrow{f} 0 \xrightarrow{f} 0.$

Notice for 0 iterates that the maximum size is $3 \approx 2^{(0+1)}$; for 1 iterate the maximum size is $5 \approx 2^{(1+1)}$; and for 2 iterates the maximum size is $8 = 2^{(2+1)}$. By repeating this argument, we find that the number of elements in a maximum set increases approximately by a multiple of 2 as the number of allowed iterates of T is increased. This makes sense since the absolute value of the slope of T is 2 everywhere except at $\frac{1}{2}$. Consequently, the local effect of the map T is to expand the distance between two points by a factor of 2.

The goal is to measure how the number of elements in a maximum set increases in the long run i.e. as n, the number of iterates, goes to ∞. Even for a simple map like the tent map, this number grows exponentially as a function of n. To end up with a finite limit, the log of the cardinality of the maximum set is divided by the number of iterates. This motivates the following mathematical definition that originates in [BOWEN].

DEFINITION 4.16. *Suppose* (X, d) *is a compact metric space. Suppose* $f_i : X \longrightarrow X$ *is a sequence of continuous functions. We say* $S \subset X$ *is* $(0, n, \epsilon, \{f_i\})$ *separated by the sequence of functions* $\{f_i\}$ *if for any* $x, y \in S$ *where* $x \neq y$*, there exists some* j*, dependent on* x *and* y*, with* $0 \leq j < n$ *so that* $d(f_j \circ f_{j-1} \circ, \ldots, f_2 \circ f_1(x), f_j \circ f_{j-1} \circ, \ldots, f_2 \circ f_1(y)) > \epsilon$*. Notice that if* $f_i = f$ *for all* i*, then this condition reduces to the standard definition of an epsilon separated set. Set* $r_{sep}(0, n, \epsilon, f_i) = max\{|S| : S$ *is* (n, ϵ) *separated by* $f_i\}$*. Set* $h_{sep}(\epsilon, f_i) = \limsup_{n \to \infty} \frac{\log r_{sep}(0,n,\epsilon,f_i)}{n}$ *and define* $h_{sep}(f_i) = \lim_{\epsilon \to 0} h_{sep}(\epsilon, f_i)$*.*

This means that $h_{sep}(f_i)$ is a crude measure of the separation of $\log r_{sep}(0, n, \epsilon, f_i)$. The next definition, of spanning sets, is a notion that is the dual of separating set. The spanning set notion measures the minimal number of points needed so that during the first n iterates each point of the space is within ϵ of one of the iterates of a point in the spanning set. As we soon show, the use of spanning set is another way of computing the entropy of a non-autonomous system. It is convenient to have two different ways of computing the entropy when proving theorems about entropy.

DEFINITION 4.17. *A set* S*, where* $S \subset K$*,* $(0, n, \epsilon, \{f_i\})$ *spans* K *if for any* $x \in K$*, there exists* $s \in S$ *such that* $d(f_j \circ \cdots \circ f_1(x), f_j \circ \cdots \circ f_1(s)) \leq \epsilon$ *for all* j*, where*

$0 \leq j < n$. *Set* $r_{span}(0,n,\epsilon,f_i,K) = min\{|S| : S\ (0,n,\epsilon,\{f_i\})$ *spans* K *by* $f_i\}$. *Set* $h_{span}(\epsilon,f_i,K) = \limsup_{n\to\infty} \frac{\log r_{span}(0,n,\epsilon,f_i,K)}{n}$ *and set*
$h_{span}(f_i,K) = \lim_{\epsilon\to 0} h_{span}(\epsilon,f_i,K)$.

If we omit K we are assuming K equals the whole space X. We establish that there are three different yet equivalent ways to compute the entropy. (This is known for autonomous systems, [BOWEN].) The following lemma establishes the equivalence of separating set entropy and spanning set entropy.

LEMMA 4.3. $h_{sep}(f_i) = h_{span}(f_i)$

Proof: Let $E_{sep}(n,\epsilon)$ be a maximal $(0,n,\epsilon,\{f_i\})$ separated set for X and let $x \in X$. There is some $y \in E_{sep}(n,\epsilon)$ so that $d(f_j \circ \cdots \circ f_1(x), f_j \circ \cdots \circ f_1(y)) \leq \epsilon$ for all $0 \leq j < n$. This assertion follows because if $x \notin E_{sep}(n,\epsilon)$, then the set $E_{sep}(n,\epsilon) \cup \{x\}$ is a $(0,n,\epsilon,\{f_i\})$ separated set for X and thus $E_{sep}(n,\epsilon)$ would not be maximal. This conclusion contradicts that $E_{sep}(n,\epsilon)$ is a maximal set. Hence, $E_{sep}(n,\epsilon)$ also $(0,n,\epsilon,\{f_1,f_2,\dots\})$ spans X with respect to f_i. Consequently, we have that

$$r_{sep}(0,n,\epsilon,f_i) = |E_{sep}(n,\epsilon)| \geq r_{span}(0,n,\epsilon,f_i). \tag{4.1}$$

Let $E_{sep}(n,2\epsilon)$ be a maximal $(n,2\epsilon)$ separated set for X, and let $E_{span}(n,\epsilon)$ be a minimal (n,ϵ) spanning set for X. Using the fact that the set $E_{span}(n,\epsilon)$ $(0,n,\epsilon,\{f_i\})$ spans all of X we define a map $T : E_{sep}(n,2\epsilon) \longrightarrow E_{span}(n,\epsilon)$ in the following manner. Let $x \in E_{sep}(n,2\epsilon)$. By the definition of $E_{span}(n,\epsilon)$, there is some $y \in E_{span}(n, \epsilon)$ so that $d(f_j \circ \cdots \circ f_1(x), f_j \circ \cdots \circ f_1(y)) \leq \epsilon$ for any j with $0 \leq j < n$. So define $y = T(x)$.

We claim T is 1 to 1. Suppose $z = T(x_1) = T(x_2)$ for some x_1, $x_2 \in E_{sep}(n, 2\epsilon)$. By the triangle inequality, for any j with $0 \leq j < n$, we have that $d(f_j \circ \cdots \circ f_1(x_1), f_j \circ \cdots \circ f_1(x_2)) \leq d(f_j \circ \cdots \circ f_1(x_1), f_j \circ \cdots \circ f_1(z)) + d(f_j \circ \cdots \circ f_1(z), f_j \circ \cdots \circ f_1(x_2)) \leq 2\epsilon$ by the definition of $z = T(x_1) = T(x_2)$. Since $E_{sep}(n, 2\epsilon)$ is a $(0, n, 2\epsilon, \{f_i\})$ separated set, the previous inequality implies that $x_1 = x_2$.

Since T is 1 to 1, $r_{sep}(0, n, 2\epsilon, f_i) = |E_{sep}(n, 2\epsilon)| \leq |E_{span}(n, \epsilon)| = r_{span}(0, n, \epsilon, f_i)$. Hence, for any $\epsilon > 0$

$$r_{sep}(0, n, 2\epsilon, f_i) \leq r_{span}(0, n, \epsilon, f_i) \leq r_{sep}(0, n, \epsilon, f_i). \tag{4.2}$$

Since $h_{span}(\epsilon, f_i) = \limsup_{n \to \infty} \frac{\log r_{span}(0,n,\epsilon,f_i)}{n}$, we have that $h_{sep}(\epsilon, f_i) = \limsup_{n \to \infty} \frac{\log r_{sep}(0,n,\epsilon,f_i)}{n}$, and we obtain

$$h_{sep}(2\epsilon, f_i) \leq h_{span}(\epsilon, f_i) \leq h_{sep}(\epsilon, f_i). \tag{4.3}$$

If $\lim_{\epsilon \to 0} h_{sep}(\epsilon, f_i)$ is finite, then elementary theorems about limits applied to equation 4.3 as $\epsilon \to 0$ finishes the proof; otherwise, if $h_{sep}(2\epsilon, f_i)$ is infinite for small enough ϵ values, equation 4.3 implies that $h_{sep}(\{f_i\}) = \infty$, and $h_{span}(\{f_i\}) = \infty$. ■

Our next goal is to introduce a way to compute the entropy of a continuous function even for spaces that may not have a metric. We only assume that X is a compact topological space. When X is a metric space, then the topological measure of entropy agrees with the one defined in terms of ϵ spanning or ϵ separating sets. Presenting entropy in terms of open covers originated from [ADLER].

Definition 4.18. *Suppose X is compact. For any open cover $\mathfrak{U}$ of X, let $N(\mathfrak{U})$ denote the number of sets in a subcover of minimal cardinality. (A subcover is minimal if no subcover contains fewer members.) Define the entropy of $\mathfrak{U}$ to be* $H(\mathfrak{U}) = \log N(\mathfrak{U})$.

Definition 4.19. *For any two open covers $\mathfrak{U}, \beta$ of X, define $\mathfrak{U} \vee \beta = \{A \cap B : A \in \mathfrak{U},\ B \in \beta\}$. Note that $\mathfrak{U} \vee \beta$ is also an open cover of X.*

The following are some properties of open covers proved in [ADLER].

Property 4.1. $N(\mathfrak{U} \vee \beta) \leq N(\mathfrak{U})N(\beta)$ *and* $H(\mathfrak{U} \vee \beta) \leq H(\mathfrak{U}) + H(\beta)$.

Property 4.2. *If $\phi : X \longrightarrow X$ is continuous, then* $\phi^{-1}(\mathfrak{U} \vee \beta) = \phi^{-1}(\mathfrak{U}) \vee \phi^{-1}(\beta)$.

Property 4.3. *If ϕ is continuous, the minimal number of elements in an open cover $N(\mathfrak{U})$, is always greater than or equal to the number of elements in the inverse image of an open cover; i.e.* $N(\mathfrak{U}) \geq N(\phi^{-1}(\mathfrak{U}))$.

Note: when ϕ is onto, then $N(\mathfrak{U}) = N(\phi^{-1}(\mathfrak{U}))$.

The next definition is analogous to Definition 13. The elements of the open cover roughly correspond to ϵ balls, except the open sets do not all have the same diameter.

DEFINITION 4.20. *Suppose $f_i : X \longrightarrow X$ is a sequence of continuous functions. Define $h_{cov}(f_i, \mathfrak{U}) =$*

$$\limsup_{n\to\infty} \frac{H[\mathfrak{U} \vee f_1^{-1}(\mathfrak{U}) \vee (f_2 \circ f_1)^{-1}(\mathfrak{U}) \vee \cdots \vee (f_{n-1} \circ f_{n-2} \ldots f_2 \circ f_1)^{-1}(\mathfrak{U})]}{n}.$$

DEFINITION 4.21. *Define the open cover entropy of $(X, \{f_i\})$ to be $h_{cov}(f_i) = \sup\{h(f_i, \mathfrak{U}) : \mathfrak{U}$ is an open cover of $X\}$.*

REMARK 4.8. *Notation: we set*

$$N(\mathfrak{U}, f_i, n) = N[\mathfrak{U} \vee f_1^{-1}(\mathfrak{U}) \vee (f_2 \circ f_1)^{-1}(\mathfrak{U}) \vee \cdots \vee (f_{n-1} \circ f_{n-2} \ldots f_2 \circ f_1)^{-1}(\mathfrak{U})],$$

$$H(\mathfrak{U}, f_i, n) = H[\mathfrak{U} \vee f_1^{-1}(\mathfrak{U}) \vee (f_2 \circ f_1)^{-1}(\mathfrak{U}) \vee \cdots \vee (f_{n-1} \circ f_{n-2} \ldots f_2 \circ f_1)^{-1}(\mathfrak{U})].$$

We now show the open cover entropy is equal to the spanning set entropy. This shows that the notions of open cover entropy, spanning set entropy, and separating set entropy are equivalent for non-autonomous systems.

LEMMA 4.4. *For any $\epsilon > 0$ and any n we can find an open cover $\mathfrak{U}$ of X so that $r_{span}(0, n, \epsilon, f_i) \leq N(\mathfrak{U}, f_i, n)$*

Proof: Fix n as a natural number. Fix $\epsilon > 0$. Since X is compact, the functions $f_1, f_2 \circ f_1, \ldots, f_{n-1} \circ \cdots \circ f_1$ are uniformly continuous on X, and since there are a finite number of functions, we can find an open cover $\mathfrak{U}$ of X so that the following $n - 1$ conditions hold:

$$U \in \mathfrak{U} \Rightarrow diam(U) < \epsilon, \tag{4.1}$$

$$U \in f^{-1}(\mathfrak{U}) \Rightarrow diam(U) < \epsilon, \tag{4.2}$$

$$U \in (f_n \circ \dots f_1)^{-1}(\mathfrak{U}) \Rightarrow diam(U) < \epsilon. \tag{4.3}$$

Let V be an element of the open cover $\mathfrak{U} \vee f_1^{-1}(\mathfrak{U}) \vee (f_2 \circ f_1)^{-1}(\mathfrak{U}) \vee \cdots \vee (f_{n-1} \circ f_{n-2} \dots f_2 \circ f_1)^{-1}(\mathfrak{U})$. Since $V = U_1 \cap f_1^{-1}(U_2) \cap (f_2 \circ f_1)^{-1}(U_3) \cap \cdots \cap (f_{n-1} \circ f_{n-2} \dots f_2 \circ f_1)^{-1}(U_n)$ and $U_1 \in \mathfrak{U}$, this implies that diam$(V) \leq diam(U_1) < \epsilon$. Hence, there is an $x \in V$ so that

$$V \subset B(x, \epsilon). \tag{4.4}$$

Let $V_1, \dots, V_m$ be a minimal subcover of $\mathfrak{U} \vee f_1^{-1}(\mathfrak{U}) \vee (f_2 \circ f_1)^{-1}(\mathfrak{U}) \vee \cdots \vee (f_{n-1} \circ f_{n-2} \dots f_2 \circ f_1)^{-1}(\mathfrak{U})$. By equation 4.4, there exists a set $\{x_1, \dots, x_m\}$ so that $V_i \subset B(x_i, \epsilon)$ and $x_i \in V_i$.

We claim our final step shows that the set $\{x_1, \dots, x_m\}$ $(0, n, \epsilon, \{f_i\})$ spans X.

Let $y \in X$. Since $V_1, \dots, V_m$ cover X. W.L.O.G., suppose $y \in V_1 = V_{01} \cap f_1^{-1}(V_{11}) \cap (f_2 \circ f_1)^{-1}(V_{21}) \cap \dots (f_{n-1} \circ f_{n-2} \dots f_2 \circ f_1)^{-1}(V_{(n-1)1})$. Because $V_1 \subset B(x_1, \epsilon)$ and $x_1 \in V_1$ we conclude that $d(x_1, y) < \epsilon$. Further, $x_1, y \in V_1 \subset f_1^{-1}(V_{11})$ implies that $f_1(x_1), f_1(y) \in V_{11}$. By the second of the n conditions $diam(V_{11}) < \epsilon$, so $d(f_1(x_1), f_1(y)) < \epsilon$. Similarly, for j with $1 \leq j < n$, we have that

$f_j \circ \cdots \circ f_1(x), f_j \circ \cdots \circ f_1(y) \in V_{j1}$, so $d(f_j \circ \cdots \circ f_1(x), f_j \circ \cdots \circ f_1(y)) < \epsilon$. Hence, $d_{n,f_i}(x_1, y) < \epsilon$. Thus, $r_{span}(0, n, \epsilon, f_i) \leq N(\mathfrak{U}, f_i, n)$. ■

The following Corollary enables us to show that the spanning set entropy is less than or equal to the open cover entropy.

COROLLARY 4.1. *For any $\epsilon > 0$, there exists an open cover $\mathfrak{U}$ of X so that*

$$\limsup_{n\to\infty} \frac{\log(r_{span}(0, n, \epsilon, f_i))}{n} \leq \limsup_{n\to\infty} \frac{H(\mathfrak{U}, f_i, n)}{n}$$

Proof: Notice that $\log(x)$ is strictly increasing. The rest of the proof follows easily from the previous lemma. ■

COROLLARY 4.2. *The spanning set entropy is less than or equal to the open cover entropy; i.e. $h_{span}(f_i) \leq h_{cover}(f_i)$.*

Proof: This is an immediate conclusion from the definitions of spanning set entropy and open cover entropy and Corollary 4.1. ■

We have shown that the spanning set entropy is less than or equal to the open cover entropy. Next we show that the spanning set entropy is greater than or equal to the open cover entropy.

LEMMA 4.5. *Fix a natural number, n. For any open cover $\mathfrak{U}$, there exists $\epsilon > 0$ so that $r_{span}(0, n, \epsilon, f_i) \geq N(\mathfrak{U}, f_i, n)$.*

Proof: Set $m = N(\mathfrak{U}, f_i, n)$. Let $V_1, \ldots, V_m$ be a minimal subcover of

$\mathfrak{U} \vee f_1^{-1}(\mathfrak{U}) \vee (f_2 \circ f_1)^{-1}(\mathfrak{U}) \vee \cdots \vee (f_{n-1} \circ f_{n-2} \ldots f_2 \circ f_1)^{-1}(\mathfrak{U})$. Since $V_1, \ldots, V_m$ is a minimal subcover of X, for each V_i there exists $\epsilon_i > 0$ and $x_i \in V_i$ so that $B(\epsilon_i, x_i) \subset V_i$ and $B(\epsilon_i, x_i) \cap V_j = \varnothing$ when $j \neq i$. Set $\epsilon = \frac{1}{2}\min\{\epsilon_1, \ldots, \epsilon_m\}$. Then the points $\{x_1, \ldots, x_m\}$ are an $(0, n, 2\epsilon, \{f_i\})$ separated set. Hence, $r_{sep}(0, n, 2\epsilon, f_i) \geq m = N(\mathfrak{U}, f_i, n)$. By Lemma 4.3, $r_{span}(0, n, \epsilon, f_i) \geq r_{sep}(0, n, 2\epsilon, f_i)$. Hence, we found an $\epsilon > 0$ so that $r_{span}(0, n, \epsilon, f_i) \geq N(\mathfrak{U}, f_i, n)$ ■

COROLLARY 4.3. *For any open cover $\mathfrak{U}$ of X, there exists $\epsilon > 0$ so that*

$$\limsup_{n\to\infty} \frac{\log(r_{span}(0, n, \epsilon, f_i))}{n} \geq \limsup_{n\to\infty} \frac{H(\mathfrak{U}, f_i, n)}{n}$$

Proof: Again, the proof is that $\log(x)$ is strictly increasing. Now, apply Lemma 4.5. ■

COROLLARY 4.4. *For a metric space, the separating set, spanning set, and open cover entropies are equivalent, i.e. $h_{sep}(f_i) = h_{span}(f_i) = h_{cover}(f_i)$*

Proof: The first equality is Lemma 4.3. The second equality follows from Corollary 4.2 and Corollary 4.3. ■

The previous Corollary suggests that we ought to use the same notation for the three different ways of defining entropy; from now on, we write $h(f_i)$ to represent topological entropy and drop all subscripts.

Now we turn our attention to an extension of Bowen's results that the topological entropy of the non-autonomous system $(X, \{f_1, \ldots, f_r\})$ with period r is equal to the topological entropy of $\{f_1, \ldots, f_r\}$ restricted to the non-wandering points.

Recall the definition of non-wandering point in Definition 1.7. Notice that this definition is logically equivalent to: For any neighborhood U of p there exists $k > 0$ such that $f_k \circ \cdots \circ f_3 \circ f_2 \circ f_1(U) \cap U \neq \emptyset$.

The next Theorem states that the entropy of a non-autonomous dynamical system with period 2, restricted to the non-wandering set is the same as the entropy of the dynamical system on the whole space. While reading the proof, notice that none of the steps rely on the fact that the integer m is a multiple of 2; these steps only depend on the fact that m is finite. Consequently, for a non-autonomous dynamical system with period p, $(X, \{g_1, g_2, \ldots, g_p, g_1, g_2, \ldots, g_p, \ldots\})$, we can set $m = kp$, and make precisely the same arguments as the proof for period 2. In the interests of clarity, we show the proof for period 2, rather than for period p.

THEOREM 4.13. *The topological entropy of $\{f, g\}$ on X is equal to the topological entropy of $\{f, g\}$ on Ω.*

Proof: Let $\epsilon > 0$ and fix m where m is even and $m \geq 1$. Let $E_{span}(0, m, \epsilon, \Omega)$ be a minimal $(0, m, \epsilon, \{f, g\})$ spanning set. Set $U = \{x \in X : d_{(0,m)}(x, y) < \epsilon$ for some $y \in E_{span}(0, m, \epsilon, \Omega)\}$. Notice that U is an open set because

$$d_{(0,m)}(z, y) \leq d_{(0,m)}(z, x) + d_{(0,m)}(x, y) \tag{4.5}$$

Choose $\delta > 0$ so that $d(z, x) < \delta$ implies that $d_{(0,m)}(z, x) < \epsilon - d_{(0,m)}(x, y)$. Hence, the inequality $d_{(0,m)}(z, y) < \epsilon$ implies that $B(x, \delta) \subset U$.

Let $X \backslash U$ denote the set theoretic complement of U in X. Then $X \backslash U$ is compact because X is compact by assumption. The next Remark essentially tells us that

$X \backslash U$ is a subset of the wandering points (the set theoretic complement of the non-wandering points). However, to make the proof more readable we present the proof of this Remark at the end of the proof of the Theorem.

REMARK 4.9. *There exists some $\beta > 0$ so that $\beta \leq \epsilon$ and for any $y \in X \backslash U$ we have $[g, f]^k(B(y, \beta)) \cap B(y, \beta) = \emptyset$ for all $k \geq 1$.*

The key idea here is to split the whole space X into two pieces, $X \backslash U$, and Ω, the non-wandering set. Then we show that the size of the spanning sets of $X \backslash U$ are bounded above by a polynomial, so $X \backslash U$ contributes nothing to the entropy.

Let $S = E_{span}(0, m, \beta, X \backslash U)$ be a minimal $(0, m, \beta, \{f, g\})$ spanning set, for $X \backslash U$, and let $E_{span}(0, m, \epsilon, \Omega)$ be a minimal $(0, m, \epsilon, \{f, g\})$ spanning set for Ω. Hence, $|S| = r_{span}(0, m, \beta, \{f, g\}, X \backslash U)$. Set $G_{span}(m) = E_{span}(0, m, \epsilon, \Omega) \cup E_{span}(0, m, \beta, X \backslash U)$. Notice that $G_{span}(m)$ is a $(0, m, \epsilon, \{f, g\})$ spanning set for X. Hence, $|G_{span}(m)| \geq r_{span}(0, m, \epsilon, \{f, g\}, X)$. Keep in mind, that our long term goal is to find an upper bound for $r_{sep}(0, n, 2\epsilon, X, \{f, g\})$.

Let $l \in \mathbb{N}$. Define $\Phi_l : X \longrightarrow G_{span}(m)^l$ where $\Phi_l(x) = (y_0, \ldots, y_l)$, and we choose y_i in the following way. If $[g, f]^{im}(x) \in U$, then choose a y_i so that $y_i \in E_{span}(0, m, \epsilon, \Omega)$ and $d_{(0,m)}([g, f]^{im}(x), y_i) < \epsilon$. Otherwise, $[g, f]^{im}(x) \in X \backslash U$, so choose a y_i so that $y_i \in E_{span}(0, m, \beta, X \backslash U)$ and $d_{(0,m)}([g, f]^{im}(x), y_i) < \beta$. Since $E_{span}(0, m, \epsilon, \Omega)$ is an $(0, m, \epsilon, \{f, g\})$ spanning set for U and $E_{span}(0, m, \beta, X \backslash U)$ is an $(0, m, \beta, \{f, g\})$ spanning set for $X \backslash U$, it is possible to make these choices of y_i to define Φ_l.