Chapter II. Topological Dynamics.

§4. Basic Concepts.[1]

By definition, *topological dynamics* is the study of those properties of dynamical systems which are invariant under topological conjugacy. This preliminary section will describe some of the basic concepts and easy results from this field.

Convention: All topological spaces are assumed to be Hausdorff. In fact we will usually assume that our spaces are metric[2] and locally compact, or even compact, with infinitely many elements.

A general *topological dynamical system* is sometimes defined as a pair $(X\,,\Gamma)$ consisting of a topological space X together with a group or semi-group Γ of continuous transformations from X to itself. However, we are interested only in dynamical systems which model the evolution of some physical system with time. Hence, for our purposes, Γ will always be either the additive group of real numbers $\mathbb{R}$, or some additive subset. In fact there are four important cases, according as Γ is equal to the set of integers $\mathbb{Z}$, the non-negative integers $\mathbb{N}$, the real numbers $\mathbb{R}$, or the non-negative real numbers $\mathbb{R}_+$.

Case 1. $\Gamma = \mathbb{Z}$**.** Given any homeomorphism $f : X \to X$, we can consider the group consisting of all iterates $f^{\circ n}$ where n can be any integer: positive, negative, or zero.

Case 2. $\Gamma = \mathbb{N}$**.** If $f : X \to X$ is only required to be a continuous map, we can still consider the semi-group consisting of all *forward iterates* $f^{\circ n}$, with $n \ge 0$. This is the case that will be emphasized in these notes. Cases 1 and 2 are quite similar to each other, but there are subtle differences. (Compare §4E.) In both of these cases, we speak of a dynamical system *with discrete time*. We will usually use the brief notation (X, f) for such systems, in place of the more formal notation $(X\,,\{f^{\circ n}\}_{n\in\mathbb{Z}})$ or $(X\,,\{f^{\circ n}\}_{n\ge 0})$.

Recall from §2A that two such dynamical systems $(X\,,f)$ and $(Y\,,g)$ are said to be *topologically conjugate* (or *topologically isomorphic*) if there is a homeomorphism h from X onto Y which satisfies the identity $g\circ h = h\circ f$. We also say that h topologically *conjugates* f to g. By definition, the "topological" properties of the dynamical system $(X\,,f)$ are those which are preserved by such a topological conjugacy.

In many interesting applications, the space X is provided with some additional geometric structure. For example it may be a smooth manifold, or may have a preferred volume element or measure. However, such non-topological properties may well be destroyed by a topological conjugacy. Topological dynamics can be described as that part of dynamics which can be formulated purely in terms of topology, and does not involve any such additional structure.

Case 3. $\Gamma = \mathbb{R}$**.** The case of a group of transformations parametrized by the real numbers is also of fundamental importance, although it will not be emphasized in these notes. Consider solutions to a differential equation of the form

$$dx/dt \;=\; v(x)\;,$$

[1] Draft from 1995, revised September 2001. I am indebted to S. Zakeri for his help with this section.

[2] One might be tempted to refer to "metric dynamics" when discussing properties (such as sensitive dependence) which depend on a specific choice of metric. Unfortunately this would be confusing since many authors, going back at least to [Rochlin], use "metric" as an abbreviation for "measure theoretic".

where x varies over a smooth manifold M and $v(x)$ is a smooth vector field on M. The solution curve $x = x(t)$ with initial condition $x(0) = x_0$ will be written as $t \mapsto f_t(x_0)$. If the manifold M is compact, or more generally whenever these solution curves exist for all $x_0 \in M$ and all $t \in \mathbb{R}$, we obtain a *one-parameter group of diffeomorphisms* $\{f_t : M \to M\}$, or briefly a *flow*, where $f_0 = \text{identity map}$ and

$$f_s \circ f_t = f_{s+t} \quad \text{for all} \quad s\,, t \in \mathbb{R}\,. \tag{4:1}$$

There is a corresponding concept of topological conjugacy for flows.

Case 4. $\Gamma = \mathbb{R}_+$ **.** If the solutions are defined only for $t \ge 0$, then we obtain a *semi-flow* or *one-parameter semi-group* of mappings from M into itself. This case occurs, for example, when M is a compact manifold with boundary, and v is a smooth vector field which points in (or at least does not point out) at all boundary points. Another important case occurs in the study of the *heat equation*, for example $\partial u/\partial t = \sum \partial^2 u/\partial x_j^2$ where $(x_1, \ldots, x_n)$ varies over a torus $\mathbb{R}^n/\mathbb{Z}^n$. Here u varies over the "infinite dimensional manifold" M consisting of all smooth real or complex valued functions on $\mathbb{R}^n/\mathbb{Z}^n$. The solutions $u(t, x_1, \ldots, x_n)$ are smooth functions, which usually can be defined only for $t \ge 0$.

In either the case of a flow or a semi-flow, we will say briefly that $(M\,, \{f_t\})$ is a dynamical system with *continuous time*. Much of the theory for continuous time is quite similar to the theory for discrete time, and we will usually skip details. However, there are some significant differences. (See for example Problem 4-b.)

If we start with a system with continuous time t, but specialize to non-negative integer values of t, then evidently $\{f_t\}_{t=1,2,\cdots}$ is just the semi-group consisting of all iterates of the *time one map* $f_1 : M \to M$. For other relations between continuous and discrete time, see §1C.

Remark. One important generalization, which will not be discussed here, concerns discontinuous maps. (Compare §3C.) Another concerns maps $f : U \to X$ which are not defined everywhere in X, but only on some subset, usually an open subset. Important examples are the theory of iterated meromorphic functions $f : \mathbf{C} \to \mathbf{C} \cup \infty$ (see [Bergweiler]), and the theory of iterated polynomial-like mappings [Douady and Hubbard 1985]. Similarly one can study a flow $\{f_t : U_t \to X\}$ which is only defined on some open set which may become smaller as $t \to \infty$. For example in celestial mechanics (§1A) one needs to exclude initial conditions which lead to collisions or other singularities. (See [Xia].)

---

§**4A. Periodicity and Limiting Behavior.** One of the simplest examples of a property which is invariant under topological conjugacy of a dynamical system (X, f) is the existence or non-existence of *periodic orbits*. Given any point $x_0 \in X$, let

$$f \;:\; x_0 \;\mapsto\; x_1 \;\mapsto\; x_2 \;\mapsto\; \cdots$$

be the *forward orbit* of x_0 under f, where $x_k = f^{\circ k}(x_0)$. By definition, the point x_0 is *periodic* under f if these successive iterates satisfy the condition that $x_0 = x_n$ for some integer $n \ge 1$, and hence $x_i = x_{n+i}$ for all i. The smallest such n is called the *period* of x_0. If h conjugates f to g, note that

$$g \;:\; h(x_0) \;\mapsto\; h(x_1) \;\mapsto\; h(x_2) \;\mapsto\; \cdots$$

is then a forward orbit under g . Thus, if x_0 is periodic of period n under f , then $h(x_0)$ is periodic of period n under g .

Every periodic point for f can also be described as a fixed point for some iterate $f^{\circ n}$. Note however that the fixed points of $f^{\circ n}$ include not only points of period n under f but also points whose period is any proper divisor of n .

Examples. An irrational rotation of the circle evidently has no periodic points at all. The doubling map m_2 of §2B is more interesting. Its only fixed point, $2\tau \equiv \tau \pmod{\mathbb{Z}}$ is the zero point, $\tau \equiv 0 \pmod{\mathbb{Z}}$. However, for the n-fold iterate the corresponding congruence

$$2^n \tau \;\equiv\; \tau \qquad (\mathrm{mod}\ \mathbb{Z})$$

has $2^n - 1$ distinct solutions $\tau \equiv k/(2^n - 1)$ modulo $\mathbb{Z}$. One of these $2^n - 1$ fixed points of $m_2^{\circ n}$ is just the fixed point $\tau \equiv 0$ of m_2 itself. If n can be expressed non-trivially as a product, then some of these fixed points of $m_2^{\circ n}$ actually have period n_1 under f , where $1 < n_1 < n$. However, if n is a prime number, then every $k/(2^n - 1)$ with $0 < k < 2^n - 1$ must actually be a periodic point of period exactly n . *Thus, if $p > 1$ is prime, the doubling map m_2 has exactly $(2^p - 2)/p$ distinct periodic orbits of period p .* There is one orbit $\{1/3\,,\,2/3\}$ of period $p = 2$, two orbits $\{1/7\,,\,2/7\,,\,4/7\}$ and $\{3/7\,,\,6/7\,,\,5/7\}$ of period $p = 3$, and so on. We will develop these ideas further in §6.

So far, we have not actually made use of the topology of X . However, the following definitely does make use of this topology.

Definition. For any $x_0 \in X$, the collection of all accumulation points for the forward orbit $\{x_0\,,\,x_1\,,\,x_2\,,\,\ldots\}$ is called the *ω-limit set* $\omega(x_0\,,\,f) \subset X$. A point $y \in X$ belongs to this ω-limit set if and only if, for every neighborhood N of y , there exist arbitrarily large integers k so that $x_k \in N$. If X is a metric space, this is clearly equivalent to the condition that there exists an infinite subsequence of $\{x_k\}$, indexed by integers $k_1 < k_2 < \cdots$, so that the distance $\mathbf{d}(x_{k_i}\,,\,y)$ tends to zero as $i \to \infty$.

This ω-limit set $\omega(x_0\,,\,f)$ is always a closed subset of X . For if y is *not* in $\omega(x_0\,,\,f)$, then there exists a neighborhood N of y so that $x_k \notin N$ for large k , and it follows that the entire neighborhood N is disjoint from $\omega(x_0\,,\,f)$. Note also that: *An ω-limit set is always forward sub-invariant, that is $f(\omega(x_0\,,\,f)) \subset \omega(x_0\,,\,f)$.* Thus, if $X' = \omega(x_0\,,\,f) \subset X$, then the pair $(X'\,,\,f|X')$ can be considered as a dynamical system in its own right. The proof is easily supplied.

In a non-compact space, this ω-limit set may well be vacuous. For example, if $f(x) = x + 1$ for $x \in \mathbb{R}$, then $\omega(x\,,\,f)$ is certainly vacuous. However:

Lemma 4.1 *If the space X is compact, then every ω-limit set $\omega(x_0\,,\,f)$ is compact, non-vacuous, and forward invariant,*

$$f(\omega(x_0\,,\,f)) \;=\; \omega(x_0\,,\,f)\ .$$

Proof. Since any infinite sequence in a compact space has at least one accumulation point, the set $\omega(x_0\,,\,f)$ is certainly non-vacuous. If the sequence $\{x_{k_i}\}$ converges to $y \in \omega(x_0\,,\,f)$, then any accumulation point y' for $\{x_{k_i - 1}\}$ will be in this ω-limit set and satisfy $f(y') = y$. $\square$

As examples, if R_α is an irrational rotation of the circle, then for any initial point τ_0 the ω-limit set $\omega(\tau_0\,,\,R_\alpha)$ is the entire circle. (Compare 3.7.) For the doubling map m_2 on the circle, $\omega(\tau_0\,,\,m_2)$ is the entire circle for Lebesgue almost every choice of starting point τ_0 . (Compare 3.8.) However, this is certainly not true for every τ_0 . For example, if τ_0 is a rational number (modulo $\mathbb{Z}$), then clearly $\omega(\tau_0\,,\,m_2)$ is finite, consisting of a single periodic orbit. Much more complicated ω-limit sets can also occur. Compare Problem 4-c.

---

§4B. Recurrence and Wandering. A point $x_0 \in X$ is said to be *recurrent* if $x_0 \in \omega(x_0\,,\,f)$, or in other words if the forward orbit of x_0 returns to an arbitrarily small neighborhood of x_0 infinitely often. As an example, every periodic point is certainly recurrent. If X is compact metric, then we will see in 4.7 that recurrent points always exist.

However, the set of all recurrent points is awkward to work with, since it may not be a closed set. As an example, consider the doubling map $t \mapsto 2t$ on the circle $\mathbb{R}/\mathbb{Z}$. Every rational number of the form $k/(2^n - 1)$ maps to itself under n-fold iteration of this map, and hence is recurrent. Thus the closure of the set of recurrent points is the entire circle. Yet the point $1/2$ with orbit $1/2 \mapsto 0 \mapsto 0 \mapsto \cdots$ is clearly not recurrent. We can get around this difficulty by passing to a slightly weaker concept:

Definition. A point $x_0 \in X$ is *non-wandering* if and only if, for every neighborhood N of x_0 , there exists an integer $k \geq 1$ so that $N \cap f^{\circ k}(N) \neq \emptyset$ (or equivalently so that $f^{-k}(N) \cap N \neq \emptyset$). In other words, there must be points arbitrarily close to x_0 which come back arbitrarily close to x_0 under some iterate of the mapping. In some sense, all of the "interesting" dynamics of the map f is concentrated in the set $\Omega = \Omega(f)$ consisting of all non-wandering points.

Conversely, x_0 is a *wandering point* if there exists a neighborhood N of x_0 which is disjoint from all of its forward images $f^{\circ k}(N)$ with $k \geq 1$. A completely equivalent statement is that the iterated pre-images $f^{-k}(N) = \{y \in X : f^{\circ k}(y) \in N\}$ are pairwise disjoint for $k \geq 0$. Evidently the subset of X consisting of all wandering points is open. It follows easily that: *The complementary set $\Omega(f) \subset X$ consisting of all non-wandering points is closed and forward sub-invariant, $f(\Omega) \subset \Omega$.*

Note that every periodic orbit is contained in this non-wandering set $\Omega(f)$, hence the closure of the set of periodic points is contained in $\Omega(f)$. However, the example of an irrational rotation of the circle shows that $\Omega(f)$ may be strictly larger than this closure. Note also that: *Every ω-limit set $\omega(x_0\,,\,f)$ is contained in $\Omega(f)$.* The proof is easily supplied. Therefore, using 4.1, we have the following.

Lemma 4.2. *If X is compact, then the non-wandering set $\Omega(f)$ is necessarily compact and non-vacuous.*

Remark. If f is actually a homeomorphism from X onto itself, then every wandering point x_0 has a neighborhood N for which *all* of the forward and backward images $f^{\circ k}(N)$ with $k \in \mathbb{Z}$ are pairwise disjoint. In particular, it follows easily that the non-wandering set is *fully invariant*:

$$f(\Omega) \;=\; f^{-1}(\Omega) \;=\; \Omega\,.$$

However, if f is not one-to-one, then these statements are certainly false. As an example, if $f(x) = a\,x\,(1-x)$ mapping the unit interval into itself, with $0 \leq a < 4$, then $+1$ is a wandering point, but its image $f(1) = 0$ is a fixed point, hence non-wandering.

For many purposes, an even weaker form of recurrence is useful. Consider a fixed mapping $f : X \to X$, where X is a metric space with distance function $\mathbf{d}(x, y)$. Given $\epsilon > 0$, by an ϵ-*pseudo-orbit*, or briefly an ϵ-*chain*, of length n from x to y we will mean a sequence of points

$$x = x_0 \ , \ x_1 \ , \ \ldots \ , \ x_n = y$$

in X , with $n \geq 1$, satisfying the condition that

$$\mathbf{d}(f(x_i)\,,\,x_{i+1}) \ < \ \epsilon \qquad \text{for} \qquad 0 \leq i < n \ .$$

(Compare Problem 3-c.) The point x is called *chain recurrent* if and only if there exists an ϵ-chain from x to itself for every $\epsilon > 0$. Intuitively, a point is chain recurrent if it can be made recurrent by an arbitrarily small perturbation of the mapping. This is certainly an appropriate concept for any empirically described mapping, which can never be known precisely. Similarly, it is appropriate for computer experiments, which usually introduce some small round-off error into any floating point computation.

It is easy to check that

$$\textbf{recurrent} \quad \Rightarrow \quad \textbf{non–wandering} \quad \Rightarrow \quad \textbf{chain recurrent} \ .$$

For example if x is non-wandering, then for any $\epsilon > 0$ we can choose a neighborhood N contained in the ϵ-neighborhood of x so that $f(N)$ is contained in the ϵ-neighborhood of $f(x)$. Since x is non-wandering, we can choose $k > 0$ so that $N \cap f^{-k}(N)$ contains at least one point y_0 . Let $y_0 \mapsto y_1 \mapsto \cdots$ be its orbit. Then x , y_1 , y_2 , $\ldots$, y_{k-1} , x is the required ϵ-chain from x to itself.

Neither of these implications can be reversed. As noted earlier, under the doubling map on the circle $\mathbb{R}/\mathbb{Z}$ the point $(1/2 \text{ modulo } \mathbb{Z})$ is non-wandering, but is not recurrent. For the map

$$f(\tau) \ = \ \tau + \sin^2(\pi\tau)/10 \qquad (4:2)$$

from $\mathbb{R}/\mathbb{Z}$ to itself, as shown (on the unit interval) in Figure 21, the non-wandering set for f is just the single point zero (modulo $\mathbb{Z}$), yet every point of the circle is chain recurrent. (See Problem 4-a. A completely equivalent example is given by the map $x \mapsto x + 1$ on the real projective line $\mathbb{R} \cup \infty$.)

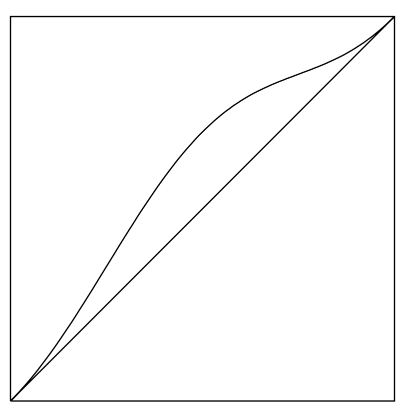

Figure 21. Graph of a map of the circle $\mathbb{R}/\mathbb{Z}$ with just one non-wandering point, but with all points chain recurrent. (*Equation* $(4:2)$ *. Here the left and right endpoints of the interval are to be identified.*)

Closely related to chain recurrence is the *chain partial ordering* of the space X , which is defined as follows. We will write $x \succ y$ if and only if there exists an ϵ-chain from x to y for every $\epsilon > 0$. Thus $x \succ x$ if and only if x is chain recurrent. If $x \succ y$ and $y \succ z$, then clearly $x \succ z$. We will say that two chain recurrent points x and y are *chain equivalent* or belong to the same *chain component* if both $x \succ y$ and $y \succ x$. Compare Problem 4-a.

There are analogous definitions for the case of a dynamical system $(X\,,\,\{f_t\})$ with continuous time. Compare Problem 4-b.

---

§4C. Transitivity and Minimality. Another useful concept is the following.

Definition. The dynamical system $(X\,,\,f)$ is *topologically transitive* if for every pair U and V of non-empty open subsets of X there exists an integer $k \geq 0$ so that $f^{\circ k}(U) \cap V \neq \emptyset$ (or equivalently $U \cap f^{-k}(V) \neq \emptyset$) .

One immediate consequence is the following.

> **Lemma 4.3.** *If (X, f) is topologically transitive, and if the space X has infinitely many points, then X has no isolated points.*

(Proofs later.) To develop these ideas, we will need the following. Recall that a subset $S \subset X$ is called *nowhere dense* if its topological closure $\overline{S}$ has no interior points.

Definition. The topological space X is called a *Baire space* if it is Hausdorff, and if every countable union of nowhere dense subsets has no interior, or equivalently if every countable intersection of dense open subsets of X is dense.

The classical Theorem of René Baire says that every complete metric space, and also every locally compact space, has this property:

> **Lemma 4.4.** *If X is either locally compact or complete metric, then X is a Baire space.*

There are several common terminologies for dealing with subsets of a Baire space X . A subset $S \subset X$ is said to be *meager* if it is a countable union of nowhere dense subsets. It is said to be *residual* if it contains a countable intersection of dense open sets, or equivalently if its complement $X \smallsetminus S$ is meager. **Definition.** We will say that a property of points in X is satisfied for a *generic* point of X if it is satisfied except for points belonging to some meager set; or equivalently if is satisfied for all points belonging to some residual set. (Caution: This use for the term "generic" is fairly common in the dynamics literature. However, the reader should beware since the word is sometimes used with other meanings.)

In the original terminology of Baire, meager sets are "of the first category", while non-meager sets are "of the second category". We can think of meager sets as a kind of topological substitute for sets of measure zero. In particular, they have analogous basic properties: *Any countable union of meager sets is meager; but the whole space X is not meager.* (However, it is definitely not true, even on the real line, that meagers sets must have measure zero, nor that sets of measure zero must be meager. Problem 4-d.)

Remark. If X is a Baire space with no isolated points, then evidently every countable subset of X is meager. In particular, such a space X must be uncountable. As one example, since the set of rational numbers is countable, it follows that *a generic real number is irrational.*

By definition, a topological space X has a *countable basis* if there are countably many non-vacuous open subsets U_i so that, for every $x \in X$ and every neighborhood N of x there exists some U_i with $x \in U_i \subset N$. With these preliminaries, we can state the following.

> **Theorem 4.5.** *If X is a Baire space with a countable basis, and if $f : X \to X$ is topologically transitive, then for a generic point $x_0 \in X$ the ω-limit set $\omega(x_0, f)$ is equal to the entire space X. Conversely, for any topological space X, if there exists a single point $x_0 \in X$ with $\omega(x_0, f) = X$, then it follows that f is topologically transitive.*

Here are some easy examples. The angle doubling map m_2 on the circle is clearly topologically transitive. In fact, if $U \subset \mathbb{R}/\mathbb{Z}$ is an open interval of length ϵ, and if $2^n\epsilon > 1$, then the image $m_2^{\circ n}(U)$ must be the entire circle $\mathbb{R}/\mathbb{Z}$. It follows that there are uncountably many dense orbits. (This statement would also follow from Borel's Theorem 3.8, but the present proof is much easier.) On the other hand, the angle doubling map on the subspace $\mathbb{Q}/\mathbb{Z}$ is also topologically transitive, but has no dense orbit. This shows that it is essential to assume that X is a Baire space. The example of the map $n \mapsto n+1$ on the space $\mathbb{N} = \{0, 1, \ldots\}$ shows that not every map with a dense orbit is topologically transitive.

Proof of 4.3. If (X, f) is topologically transitive and $y \in X$ is an isolated point, then we can certainly find some $x \in X$ with $f(x) = y$. Hence there is an entire neighborhood U of x with $f(U) = y$. Since the singleton $\{y\}$ is an open set, we have $f^{\circ n}(y) \in U$ for some $n \geq 0$. Therefore y is periodic, and it follows easily that X coincides with the orbit of y. $\square$

Proof of 4.4. Let $U_1, U_2, U_3, \ldots$ be dense open subsets of the locally compact space X, and let V_1 be an arbitrary non-vacuous open set. To construct a point in the intersection $V_1 \cap U_1 \cap U_2 \cap \cdots$, we construct nonvacuous open sets V_n with $V_1 \supset V_2 \supset V_3 \supset \cdots$ so that the closure $\overline{V}_{n+1}$ is a compact subset of $U_n \cap V_n$ as follows, by induction on n. Given V_n, since U_n is dense we can choose a point $x \in U_n \cap V_n$. Since X is locally compact, we can choose a neighborhood V_{n+1} of x which is small enough so that the closure $\overline{V}_{n+1}$ is a compact subset of $U_n \cap V_n$. Now the intersection of the compact sets $\overline{V}_2 \supset \overline{V}_3 \supset \overline{V}_4 \supset \cdots$ is non-vacuous, and is contained in $V_1 \cap U_1 \cap U_2 \cap \cdots$. This completes the proof for X locally compact. The proof in the complete metric case is quite similar, and will be left to the reader. $\square$

Proof of 4.5. Let $\{U_i\}$ be a countable basis for the open sets of the Baire space X, and let $V_i = U_i \cup f^{-1}(U_i) \cup f^{-2}(U_i) \cup \cdots$. If f is topologically transitive, then this open set V_i is dense. Now a generic point $x_0 \in X$ belongs to the intersection $\bigcap_i V_i$. It follows that the forward orbit of x_0 intersects every U_i, and hence is dense. To prove that $\omega(x_0, f) = X$ we must show that the forward orbit of x_0 hits every non-vacuous open set U infinitely often. Otherwise, there would a largest n such that $x_n = f^{\circ n}(x_0)$ belongs to U. If X has no isolated points, then the set $U \smallsetminus \{x_0, x_1, \ldots, x_n\}$ is also non-vacuous, hence the orbit of x_0 must hit it, yielding a contradiction. On the other hand, if X does have an isolated point y, then topological transitivity implies that the forward orbit of y must eventually hit every non-vacuous open set. In particular, it must hit the open set $f^{-1}(y)$; hence y must be periodic. The complement of the orbit of y is then an open set, contradicting the hypothesis of topological transitivity.

Conversely, if the ω-limit set $\omega(x_0\,,\,f)$ is equal to the entire space X , then the forward orbit of x_0 must hit every non-vacuous open set infinitely often, and topological transitivity follows easily. $\square$

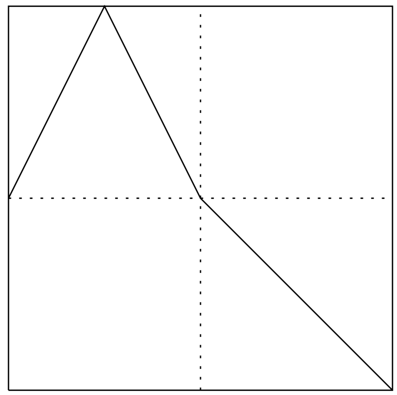

Figure 22. A topologically transitive map such that $f \circ f$ is not topologically transitive.

If f is topologically transitive, it does not follow that its iterates $f^{\circ n}$ are topologically transitive. In fact, it may happen that the space X can be decomposed into distinct pieces which are permuted cyclically by f . For example in Figure 22, f interchanges the left and right half-intervals which intersect only in a single point. However, $f^{\circ 2}$ is topologically transitive on each half-interval.

Here is a precise statement. Suppose that the space X is compact metric.

Lemma 4.6. *If $f : X \to X$ is topologically transitive but $f^{\circ n}$ is not, then there are uniquely defined compact subsets $X_i \subset X$ with $X = X_1 \cup \cdots \cup X_m$, where $m \ge 2$ is some divisor of n , so that $f(X_i) = X_{i+1}$, taking i modulo m , where $f^{\circ n}$ is topologically transitive on each X_i , and where $X_i \cap X_j$ is nowhere dense in X for $i \not\equiv j$.*

Proof. Choose a dense orbit $f : x_0 \mapsto x_1 \mapsto \cdots$, and let $X_i = X_{i+n} = \omega(x_i\,,\,f^{\circ n})$. Then $f^{\circ n}|X_i$ is topologically transitive, since the orbit of x_i is dense, and $f(X_i) = X_{i+1}$ since X is compact. If $X_i \cap X_j$ had an interior point, with $1 \le i < j \le n$, then for a generic point x in this interior we would have $\omega(x, f^{\circ n}) = X_i$ but also $\omega(x, f^{\circ n}) = X_j$, hence $X_i = X_j$. These sets X_i are uniquely defined, since a generic $x \in X$ has $\omega(x, f^{\circ n})$ equal to one of the X_i . If $m < n$ is the greatest common divisor of $j - i$ and n , then it follows easily that $X_{h+m} = X_h$ for all h . Taking the smallest possible value for m , the conclusion follows. $\square$

Some dynamical systems satisfy a condition which is much stronger than transitivity:

Definition. The dynamical system $(X\,,\,f)$ is *minimal* if *every* orbit is dense. More generally, a non-vacuous closed subset $M \subset X$ is *minimal* if, for every $x_0 \in M$, the closure of the forward orbit $\{x_0\,,\,x_1\,,\,x_2\,,\,\ldots\}$ is precisely equal to M . A completely equivalent condition, which justifies the term "minimal", is that $f(M) \subset M$ but that no closed subset $S \subset M$ with $S \ne \emptyset, M$ can satisfy $f(S) \subset S$.

As an example, every periodic orbit is certainly a minimal set. For an irrational rotation of the circle, the entire circle is a minimal set. On the other hand, for the doubling map

m_2 , the entire circle is certainly not minimal, since there are many periodic points. In fact m_2 has uncountably many distinct minimal sets, and these can be extremely complicated. (Problem 4-c.)

Note that every point x_0 in a minimal set M is necessarily recurrent. For otherwise the forward orbit of $f(x_0)$ would fail to be dense. It follows that every minimal set must be forward invariant, $M = f(M)$.

(There is an analogous concept of minimality for a one parameter group or semi-group of maps $f_t : X \to X$. One very interesting example is provided by the *horocycle flow* on the unit tangent bundle of a compact surface of constant negative curvature. See [Marcus], or Manning's article in [Bedford, Keane and Series].)

To simplify the discussion, let us assume that X is a metric space. A set $X_0 \subset X$ will be called *sub-invariant* if $f(X_0) \subset X_0$.

Lemma 4.7. *If* $f : X \to X$ *with* X *metric, then every non-vacuous compact sub-invariant set* $X_0 \subset X$ *contains a minimal set.*

Proof. Since X_0 is compact metric, there exists a *countable basis* $\{U_i\}$ for its topology, that is a countable collection of open sets so that for any neighborhood N of any point $x \in X_0$ there exists a U_i with $x \in U_i \subset N$. Starting with the given X_0 , let us construct compact sub-invariant sets $X_0 \supset X_1 \supset \cdots$ by induction, as follows. If every orbit in X_{n-1} intersects the open set U_n , then we set $X_n = X_{n-1}$. On the other hand, if some orbit in X_{n-1} is disjoint from U_n , then let X_n be the union of all orbits in X_{n-1} which are disjoint from U_n . Then clearly the sets $X_0 \supset X_1 \supset X_2 \supset \cdots$ are compact, non-vacuous, and sub-invariant. Let M be the intersection of the X_n . Then M has these same properties. Furthermore, for each of the basic open sets U_n , either *all* orbits in M intersect U_n , or else *no* orbit in M intersects U_n . From this, it follows easily that every orbit in M is dense in M . □

As a corollary, we see that: *Every compact metric dynamical system has recurrent points.* This follows from the statement that every point of a minimal set is recurrent.

---

§4D. Chaos: Sensitive Dependence, Mixing. For this subsection, we suppose that X is a metric space, with distance function $\mathbf{d}(x, y) \geq 0$.

Although the concept of "chaotic dynamics" is of fundamental importance, there is no general agreement on a definition. Some authors use "chaos" as an abbreviation for "positive topological entropy" (§7; see for example [Li and Yorke], [MacKay and Tresser]). With this usage, the map $f(z) = z^2$ on the Riemann sphere $\mathbf{C} \cup \infty$ would be consider chaotic, even though most orbits converge to either zero or infinity. My own feeling is that a dynamical system should only be called chaotic if *most* orbits behave chaotically. Thus it would be quite reasonable to call a system chaotic if it not only has positive topological entropy but also is transitive. (Compare [Glasner and Weiss].) The very idea of chaos suggests measure theoretic ideas, and definitions often involve derivatives also (in the guise of *Liapunov exponents* or *homoclinic points*; see [Guckenheimer and Holmes]). However, there are some purely topological properties which seem intimately related to chaotic behavior. The most fundamental of these is sensitive dependence, as discussed in §1B and §2C. Let us first discuss sensitive dependence at a point.

Definition. The map $f : X \to X$ exhibits *sensitive dependence* at the point $x_0 \in X$ if there exists a number $\epsilon > 0$ with the following property: *For every neighborhood of* x_0 *, there exists a point* y_0 *in the neighborhood and an integer* $n \geq 0$ *so that* $\mathbf{d}(f^{\circ n}(x_0)\,,\,f^{\circ n}(y_0)) > \epsilon$ *.*

In other words, the two orbits $x_0 \mapsto x_1 \mapsto \cdots$ and $y_0 \mapsto y_1 \mapsto \cdots$, with initial points x_0 and y_0 in the given neighborhood, must satisfy $\mathbf{d}(x_n\,,\,y_n) > \epsilon$ for at least one $n \geq 0$. Of course this integer n must depend on the choice of neighborhood: If we start with a very small neighborhood, then it will be necessary to iterate many times in order to obtain the required separation.

We will usually assume that X is compact. It is then not difficult to show that this property is invariant under topological conjugacy: it does not depend on the particular choice of metric. However, for non-compact X it definitely does depends on the metric. As an example, if we use the Euclidean metric $\mathbf{d}(x,y) = |x-y|$ on the set of positive real numbers $x > 0$, then the doubling map $x \mapsto 2\,x$ exhibits sensitive dependence at every point. But if we use the metric $\mathbf{d}(x,y) = |\log x - \log y|$, then the doubling map is an isometry, and hence is not sensitively dependent anywhere. (Equivalently, the doubling map is topologically conjugate to the translation $y \mapsto y + \log 2$ on the real line, which is not sensitively dependent, where $y = \log x$.)

The opposite of sensitive dependence is stability, in the sense of Liapunov.

Definition. The map $f : X \to X$ is *Liapunov stable* at a point $x_0 \in X$ if, for every $\epsilon > 0$, there exists a neighborhood N of x so that

$$\mathbf{d}(f^{\circ n}(x_0)\,,\,f^{\circ n}(y_0)) \;<\; \epsilon$$

for every y_0 in N and every $n \geq 0$. This same idea can be expressed by saying that the family of iterates $\{f^{\circ n}\}_{n\geq 0}$ is *equicontinuous* at x_0. Evidently f exhibits sensitive dependence at x_0 if and only if it is not Liapunov stable at x_0.

There is a classical family of examples, studied by P. Fatou and G. Julia, which displays the contrast between these two kinds of dynamic behavior in particularly striking fashion. If f is a rational map from the Riemann sphere $\mathbf{C} \cup \infty$ to itself, then by definition the set of points of Liapunov stability (= equicontinuity of the family of iterates) is called the *Fatou set* of f, and its complement is called the *Julia set*. The Fatou set is always open, or equivalently the Julia set is always compact. Furthermore, in the non-linear case, f restricted to the Julia set always exhibits sensitive dependence. (See for example [Milnor [1999].) We will return to this subject in Example 2 below.

As in §2C, if f exhibits sensitive dependence at *every* point of X, and if further the number ϵ can be chosen as a constant, independent of x_0, then we say briefly that f *exhibits sensitive dependence*. Again, this concept depends on an explicit choice of metric in general. However, if X is compact, then it does not depend on the metric.

One extremely interesting definition has been given by [Devaney], as improved by [Banks et al.]. (See also [Glasner and Weiss].)

Definition. A dynamical system (X, f) is *chaotic in the sense of Devaney*, if X has infinitely many elements, and if the following two conditions are satisfied:

(i) the set of all periodic points is everywhere dense in X, and

(ii) f is topologically transitive.

The following statement is unusual in that the conclusion makes essential use of the metric, although the hypothesis does not.

Theorem 4.8 (Banks, Brooks, Cairns, Davis and Stacy). *If the dynamical system (X, f) is Devaney chaotic, then for any metric compatible with the topology, it exhibits sensitive dependence on initial conditions.*

Proof. Choose two different periodic orbits $\mathcal{O}_1$ and $\mathcal{O}_2$, and let $4\epsilon > 0$ be the minimum distance from a point of $\mathcal{O}_1$ to a point of $\mathcal{O}_2$. Let N be an arbitrarily small neighborhood of an arbitrary point of X. Choose a periodic point $x_0 \in N$, and let $p \geq 1$ be its period. This point x_0 must have distance $\geq 2\epsilon$ from at least one of the two preferred periodic orbits, say from $\mathcal{O}_1$. Choose a neighborhood U of a point $z_0 \in \mathcal{O}_1$ which is so small that, for $0 \leq n < p$, the image $f^{\circ n}(U)$ is contained in the open ϵ-neighborhood of the corresponding point $z_n \in \mathcal{O}_1$. Since f is topologically transitive, we can choose a point $y_0 \in N$ and an iterate $y_m = f^{\circ m}(y_0)$ which belongs to this neighborhood U. Now choose n so that $m + n$ is divisible by p, with $0 \leq n < p$. Then $x_{m+n} = f^{\circ m+n}(x_0)$ is equal to x_0, but

$$y_{m+n} \;=\; f^{\circ n}(y_m) \;\in\; f^{\circ n}(U)$$

has distance $< \epsilon$ from $z_n \in \mathcal{O}_1$, and hence has distance $> \epsilon$ from $x_{m+n} = x_0$. Thus $x_0, y_0 \in N$ but $\mathbf{d}(x_{m+n}, y_{m+n}) > \epsilon$, where ϵ is a uniform constant as required. □

One objection to this definition is that, in higher dimensions, a dynamical system may exhibit all of the other symptoms of chaotic behavior without having any periodic points at all. (Compare Example 4 below.) The following properties seem to incorporate the intuitive idea of chaotic behavior, in topological dynamics, and yet do not require any periodicity.

Definition. A dynamical system (X, f) is called *topologically 2-transitive* if the cartesian product $f \times f$ mapping $X \times X$ into itself is topologically transitive. It is *topologically mixing* if for any pair U and V of non-vacuous open sets there exists an integer n_0 so that

$$f^{\circ n}(U) \cap V \;\neq\; \emptyset$$

for *all* $n \geq n_0$. (Note: Topologically 2-transitive maps are sometimes called "weakly mixing". Compare [Denker et al.].)

Lemma 4.9. *Every topologically mixing dynamical system is topologically 2-transitive. If f is topologically 2-transitive, and if X has more than one point, then f exhibits sensitive dependence using any metric. Furthermore, if f is topologically 2-transitive and X is compact metric, then every iterate $f^{\circ n}$ is topologically transitive.*

Proof of 4.9. The first statement is clear. To prove that 2-transitivity implies sensitive dependence, choose open sets V_1 and V_2 which are separated by some positive distance. Now for any open set U we can find a forward iterate which intersects both V_1 and V_2. Finally, the statement that 2-transitivity implies that $f^{\circ n}$ is transitive follows easily from 4.6. □

In fact topological mixing is an extremely strong form of sensitive dependence. For a topologically mixing map of a compact metric space, note that $f^{\circ n}(U)$ comes within distance

ϵ of *every* point of X, provided that U is a non-vacuous open set and n is sufficiently large. *Thus, if we can only measure distances to an accuracy of ϵ, then measurement of x yields absolutely no information about the location of $f^{\circ n}(x)$ for large n.*

To conclude this section, let us give some examples.

Example 1. All of the dynamical systems studied in §2 (for example the quadratic interval map $x \mapsto 2x^2 - 1$, and the one or two-sided 2-shift) are both Devaney chaotic and topologically mixing. Proofs are easily supplied. In the case of the angle doubling map on the circle, this statement generalizes in two ways:

Example 2. Following Fatou and Julia, let f be a rational map of degree two or more from the Riemann sphere $\mathbf{C} \cup \infty$ to itself. The ***Julia set*** $J \subset \mathbf{C} \cup \infty$ can be defined as the closure of the set of repelling periodic points. Alternatively, it can be defined as the set of points which are *not* Liapunov stable. (Compare the discussion at the beginning of this subsection.) The map f restricted to its Julia set J is always Devaney chaotic and mixing. (See for example [Milnor 1999].) In the special case of the polynomial map $f(z) = z^2$, the Julia set is the unit circle, so we recover the angle doubling example.

Example 3. Now let us replace the circle $\mathbb{R}/\mathbb{Z}$ by the torus $\mathbf{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$. Consider a linear map

$$F(\tau\,,\,\upsilon) \;=\; (a\tau + b\upsilon\,,\,c\tau + d\upsilon)$$

from $\mathbf{T}^2$ to itself, where

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \tag{4:3}$$

is a matrix of integers with determinant $ad - bc \neq 0$. Thus F is a homeomorphism if $ad - bc = \pm 1$, and is a many-to-one map otherwise.

Lemma 4.10. *The periodic points of F are dense in $\mathbf{T}^2$. Furthermore, If the eigenvalues λ_1 and λ_2 of this matrix are off the unit circle, $|\lambda_j| \neq 1$, then F is topologically mixing and hence chaotic.*

Proof. For every integer $q \geq 1$ which is relatively prime to $|ad - bc|$, let A_q be the finite set consisting of all $(\tau\,,\,\upsilon) \in \mathbf{T}^2$ such that both τ and υ are rational numbers with denominator q, when expressed as fractions in lowest terms. Then it is not difficult to check that F maps A_q as a one-to-one map onto itself. In fact the union of the A_q is precisely the set of all periodic points of F. Clearly this set is dense in $\mathbf{T}^2$.

Now suppose that the eigenvalues of (4:3) are off the unit circle. If both eigenvalues are greater than 1 in absolute value, then it is not hard to see that F is locally expanding, in a suitably chosen Riemannian metric (as defined in §4E below), and hence that the successive forward iterates of any open disk contain disks which are larger and larger, until they cover the entire torus. This certainly implies topological mixing. Now suppose that $|\lambda_1| > 1 > |\lambda_2| > 0$. Since the eigenvalues are not both integers, and are not complex conjugates, they must be real and irrational. Let L be any line segment of length a which is parallel to the λ_1-eigendirection. Then the image $F(L)$ will be a parallel line segment of length $|\lambda_1|\,a > a$. Note that the slope of L is irrational, since otherwise we could choose such a line segment joining two integer points. This would imply that $F(L)$ is a parallel line joining two integer points, which is impossible since λ_1 is irrational. Any open set $U \neq \emptyset$ contains such a line segment, say of length ϵ, and the length of the n-th forward image

$F^{\circ n}(L)$ is $|\lambda_1^n|\epsilon$, which tends to infinity as $n \to \infty$. But since the slope is irrational, it is not difficult to show that these image line segments fills out the torus more and more densely as $n \to \infty$. (Compare Theorem 3.7.) In particular, for any open set $V \neq \emptyset$ we have $f^{\circ n}(L) \cap V \neq \emptyset$ for large n , which again proves topological mixing. □

Example 4. I am indebted to Philip Boyland for the observation that not every topologically mixing system possesses periodic points. Consider for example any compact Riemannian manifold M of negative curvature. The *geodesic flow* $\{g_t\}$ on the unit tangent bundle T_1M assigns to each unit tangent vector v at $x \in M$ and each real number $t > 0$ the tangent vector $g_t(v)$ at the end of the geodesic segment of length t which starts at x in the direction v . The *length spectrum* of $\{g_t\}$, that is the collection of all lengths of closed geodesics, is a closed and countable collection of real numbers. Now for any $t > 0$ which does not a rational multiple of any element from this length spectrum, the map $g_t : T_1M \to T_1M$ has no periodic points at all, and hence is not chaotic in Devaney's sense. Yet this map is topologically mixing. (See for example [Hopf], [Ornstein and Weiss], [Ballmann and Brin], [Burns]; and note that if a flow $\{f_t\}$ is mixing then each individual map f_t with $t > 0$ must also be mixing.) An even more startling example is the *horocycle flow* on a compact surface of negative curvature, which is not only mixing but also minimal. (See [Marcus].) If a flow $\{f_t\}$ is minimal, then evidently no f_t with $t > 0$ can have a periodic point. Another startling example, constructed by [Rees], is a minimal homeomorphism of the 2-torus which has positive topological entropy, and hence has sensitive dependence.

On the other hand, for maps in dimension one, it is impossible to have any form of complicated dynamics without infinitely many periodic points. (Compare [Misiurewicz and Szlenk], [Milnor and Thurston].) The same is true for diffeomorphisms in dimension two. (See [Katok].) [Kuperberg] has shown that there exist smooth flows on the 3-dimensional sphere with no periodic orbits at all. However, I don't know how complicated the associated dynamics can be.

---

§4E. Forward Expansive Maps. This section will discuss several properties which are similar to sensitive dependence but much sharper. To simplify the discussion, we consider only compact metric spaces.

Definition. A map $f : X \to X$ from a compact metric space to itself is *forward expansive* (also called *positively expansive*) if there exists a number $\epsilon > 0$ so that, for any two points $x_0 \neq y_0$ in X , there exists an integer $n \geq 0$ with $\mathbf{d}(f^{\circ n}(x_0)\,,\, f^{\circ n}(y_0)) > \epsilon$.

It is not difficult to check that this property does not depend on the particular choice of metric. However, the constant ϵ does depend of the metric, and the number of iterations n must certainly depend on the choice of x_0 and y_0 . For if x_0 and y_0 are extremely close, then it is necessary to iterate many times in order to separate them.

If X has no isolated points, then clearly every forward expansive map exhibits sensitive dependence. The converse is false. For example, for the quadratic map $q(x) = 2x^2 - 1$ of §2A, given any $\epsilon > 0$ we can choose $0 < x < \epsilon/2$. Since $f(x) = f(-x)$, it follows that the maximum separation for the orbits of x and $-x$ is given by

$$\sup_{n \geq 0} |f^{\circ n}(x) - f^{\circ n}(-x)| \;=\; 2x \;<\; \epsilon \,.$$

Thus q is not forward expansive, although q exhibits sensitive dependence by §2C.

In order to illustrate this concept, let us make a detailed study of the one-dimensional case. First consider a closed interval.

Lemma 4.11. *No map f from a closed interval I to itself can be forward expansive.*

Proof. If f is forward expansive, then it must certainly be locally one-to-one. Hence it must be strictly monotone, either increasing or decreasing. Replacing f by $f \circ f$ if necessary, we may assume that it is monotone increasing. If $f(x) \le x$ for all x, then two points within ϵ of the left hand end of I will remain ϵ-close through all forward iterations. But otherwise, if U is a connected component in the set of x with $f(x) > x$, then two points near the right hand end of U will remain ϵ-close under iteration. Thus f cannot be forward expansive. □

Lemma 4.12. *A map from the circle $\mathbb{R}/\mathbb{Z}$ to itself is forward expansive if and only if it is topologically conjugate to a linear map $f(\tau) \equiv n\,\tau \pmod{\mathbb{Z}}$, where n is some integer with $|n| \ge 2$.*

Caution: Even if f is a differentiable map, it is definitely not asserted that the topological conjugacy is differentiable. (Compare Problem 4-j.)

We will give most of the proof here. However, the argument will make use of a theorem of Poincaré which will be proved only in §8.

Proof of 4.12. It is easy to check that every such linear map is forward expansive. To prove the converse statement, we will use the following two remarks. A non-degenerate interval $I \subset \mathbb{R}/\mathbb{Z}$ will be called ***periodic*** if some forward iterate $f^{\circ k}(I)$ is contained in I, and will be called a ***wandering interval*** if the forward images $f^{\circ k}(I)$ are pairwise disjoint. Note the following.

Assertion. *No forward expansive map of the circle can have either a periodic interval or a wandering interval.*

In fact the first statement follows easily from 4.11. To prove the second, if $f^{\circ k}(I) \cap f^{\circ \ell}(I) = \emptyset$ for $0 \le k < \ell < \infty$, then the sum of the lengths of the $f^{\circ k}(I)$ is finite, hence this length tends to zero as $k \to \infty$, and it follows easily that f is not forward expansive. This proves the Assertion.

Next recall that every map f from $\mathbb{R}/\mathbb{Z}$ to itself lifts to a map $F : \mathbb{R} \to \mathbb{R}$ satisfying the identity

$$F(x+1) \;=\; F(x) + n\ ,$$

where n is an integer constant called the ***degree***. If f is forward expansive, then f is locally one-to-one, and it follows, as in the proof of 4.11, that F must be strictly monotone, either increasing or decreasing. In particular, the degree n cannot be zero.

For a monotone circle map f of degree $n = 1$, a classical theorem, essentially due to Poincaré, asserts that f is topologically conjugate to an irrational rotation, $\tau \mapsto \tau + \text{constant}$, if and only if it has no wandering interval and no periodic interval. A proof of Poincaré's theorem will be given in §8. Since a rotation clearly cannot be forward expansive, it follows that no degree one map of the circle can be forward expansive. (For an alternative proof, see 4.15 below.)

Similarly, if f has degree -1 then $f\circ f$ has degree $+1$ and hence cannot be forward expansive. (For a more elementary proof, see Problem 4-e.)

The various cases with $|n|\ge 2$ are all similar to each other. To fix ideas, let us restrict attention to the case $n=2$. Then $F(k)=F(0)+2k$ for any integer k. Thus $F(k)$ is greater than k if the integer k is close to $+\infty$, and less than k for k near $-\infty$. By the intermediate value theorem, it follows that F has at least one fixed point $x_0=F(x_0)$. Now for each dyadic fraction $\alpha=m/2^k$ we will choose a number $x_\alpha\in\mathbb{R}$ by induction on k so as to satisfy the identities

$$F(x_\alpha)=x_{2\alpha}\quad\text{and}\quad x_{\alpha+1}=x_\alpha+1\ .$$

To start the induction, set $x_m = x_0+m$ for every integer m, and note that $F(x_m)=x_{2m}$. Now if $\alpha=m/2^k$ then, assuming inductively that $x_{2\alpha}$ has already been defined, we set x_α equal to the unique solution to the equation $F(x_\alpha)=x_{2\alpha}$. It is easy to see that these numbers are well defined, with $x_{\alpha+1}=x_\alpha+1$, and $x_\alpha<x_\beta$ whenever $\alpha<\beta$.

Now let $G:\mathbb{R}\to\mathbb{R}$ be the unique monotone map which satisfies $G(x_\alpha)=\alpha$ for every dyadic fraction α. In other words, we set

$$G(x)\ =\ \inf\{\alpha\ ;\ x_\alpha>x\}\ =\ \sup\{\alpha\ ;\ x_\alpha<x\}\ .$$

It is not difficult to check that G is continuous and monotone, with

$$G(x_\alpha)=\alpha\ ,\quad G(x+1)=G(x)+1\ ,\quad\text{and}\quad G(F(x))=2\,G(x)\ .$$

Thus G induces a monotone degree one map $g:\mathbb{R}/\mathbb{Z}\to\mathbb{R}/\mathbb{Z}$ which semi-conjugates f to the doubling map $\tau\mapsto 2\tau$. We must prove that g is actually a homeomorphism. Otherwise, it would map some non-trivial interval $J\subset\mathbb{R}/\mathbb{Z}$ to a single point. Taking a maximal such interval, there are two possibilities. If the image point in $\mathbb{R}/\mathbb{Z}$ is eventually periodic under doubling, then f has a periodic interval. But if this image point is not eventually periodic, then it follows that f has a wandering interval. Since both possibilities are excluded by the Assertion above, it follows that g must be a homeomorphism, as required. □

The word "expansive" should not be confused with "expanding", which refers to a stronger property which depends sharply on the particular choice of metric. In fact the reader should beware, since the word "expanding" has been used with several slightly different meanings ([Shub 1969, 1970], [Ruelle 1978], [Gromov]). However, remarkable work of [Reddy] and [Coven and Reddy] shows that many of these definitions are essentially equivalent. (See 4.14 below.)

First a slightly weaker notion.

Definition. The map f from a compact metric space to itself is *locally distance increasing* if every point has a neighborhood U so that

$$\mathbf{d}(f(x)\,,\,f(y))\ >\ \mathbf{d}(x,y)$$

whenever $x\,,\,y\in U$.

Lemma 4.13. *If X is compact with infinitely many elements, then no homeomorphism from X to itself can be locally distance increasing.*

The following rather easy proof is due to Shub. (A somewhat sharper result was proved much earlier by Gottschalk and Hedlund, by a more difficult argument. Compare 4.15 below.)

Proof of 4.13. Suppose that f is a locally distance increasing homeomorphism of a compact metric space. It follows easily, using uniform continuity, that f^{-1} is locally distance decreasing. Hence there is an $\epsilon > 0$ so that

$$\mathbf{d}(f^{-1}(x)\,,\,f^{-1}(y)) \;<\; \mathbf{d}(x,y)$$

whenever $0 < \mathbf{d}(x,y) < \epsilon$. For any integer $n \geq 1$, let $\delta(n) \geq 0$ be the smallest real number δ such that X can be covered by n closed balls $\{x\,;\,\mathbf{d}(x\,,\,x_i) \leq \delta\}$ of radius δ. An easy compactness argument shows that such a number $\delta(n)$ exists, and that $\delta(n)$ tends to zero as $n \to \infty$. In particular, we can choose n so that $\delta(n) < \epsilon$. If X has infinitely many elements, then $\delta(n) > 0$, and applying f^{-1} we obtain a covering by n closed sets, each of which is contained in a ball of radius strictly less than $\delta(n)$, which is impossible. This contradiction completes the proof. □

As an application, consider a subset X of the circle $\mathbb{R}/\mathbb{Z}$ which maps homeomorphically onto itself under the angle doubling map. *If X is compact, then it must be finite.* This follows, since the angle doubling map is clearly locally distance increasing. (For an application of this statement in holomorphic dynamics, due to Sullivan and Douady, see [Milnor 1999, §18.8].)

Next we introduce a slightly sharper condition.

Definition. The map f on a metric space f is *locally expanding* if there is a real number $\lambda > 1$ so that every point has a neighborhood U with

$$\mathbf{d}(f(x)\,,\,f(y)) \;\geq\; \lambda\,\mathbf{d}(x,y)$$

for every $x\,,\,y \in U$. We will say that f is *locally λ-expanding* when it is useful to emphasize the precise value of this constant $\lambda > 1$.

One particularly important example is the case of a compact Riemannian manifold M, with metric $\mathbf{d}(x,y)$ equal to the length of the shortest path from x to y. Every smooth map $f : M \to M$ induces a linear map Df from the tangent vector space at x to the tangent vector space at $f(x)$. Clearly f is locally expanding if and only if

$$\|Df(v)\| \;>\; \|v\|$$

for every non-zero tangent vector v, where $\|v\|$ denotes the Riemannian norm of v.

According to [Gromov], the existence of a locally expanding map on a Riemannian manifold implies very strong restrictions on its geometry. In fact, M must be diffeomorphic to a quotient L/Γ where L is a nilpotent Lie group and Γ is a discrete group of affine transformations of L. (By definition, an "affine transformation" is a composition of automorphisms and left translations.)

Clearly every locally expanding map is locally distance increasing. Furthermore, it is not difficult to check that every locally distance increasing map on a compact metric space is forward expansive (Problem 4-i). The following converse statement has been proved by [Reddy].

Theorem 4.14. *Given any forward expansive map on a compact metrizable space, there exists a new metric, compatible with the topology, so that f is locally expanding with respect to this new metric.*

I will not try to give the proof, which depends on a delicate Metrization Lemma of A. H. Frink.

Combining 4.13 and 4.14, we have the following result, which is due to [Gottschalk and Hedlund, pp. 85-86].

> **Corollary 4.15.** *If X is a compact metric space with infinitely many elements, then no homeomorphism of X can be forward expansive.*

Thus a homeomorphism of an interesting space can never be either forward expansive or locally expanding. (However, closely related concepts are very important when studying the dynamics of homeomorphisms. The analogue of forward expansiveness is discussed in Problem 4-k. The analogue of the locally expanding property is "hyperbolicity": See for example [Robinson] or [Guckenheimer and Holmes].)

We conclude this section with a rather technical lemma about locally expanding maps. [Ruelle 1978] proved important measure theoretic properties for maps from a compact metric space to itself which are "locally expanding" (as defined above) and which satisfy one further condition, namely the conclusion of the following lemma. He used the term "expanding" for maps satisfying both of these these conditions.

> **Lemma 4.16 (Coven and Reddy).** *If f is an open and locally λ-expanding map from the compact metric space X to itself, then for $\epsilon > 0$ sufficiently small, f maps the open ϵ-neighborhood $N_\epsilon(x)$ of an arbitrary point $x \in X$ homeomorphically onto an open set which contains the neighborhood $N_{\lambda\epsilon}(f(x))$.*

Note: The condition that f is an *open* map, that is, that it maps open sets to open sets, is essential. However, in the most important application, to manifolds, this condition comes free. Brouwer's classical *Theorem on Invariance of Domain* says that every one-to-one map from an open set $U \subset \mathbb{R}^n$ into $\mathbb{R}^n$ maps U homeomorphically onto an open set. (See for example [Hurewicz and Wallman].) As a corollary, *every locally one-to-one map from a topological manifold into itself is an open map.* This statement applies also to manifolds with boundary, provided that the map carries boundary into boundary.

Proof of 4.16. Otherwise, for each n , taking $\epsilon = 1/n$, we could find two points x_n and z_n so that $\mathbf{d}(f(x_n)\,,\, z_n) < \lambda/n$ but $z_n \not\in f(N_{1/n}(x_n))$. Choose some subsequence of $\{x_n\}$ which converges to a point $\hat{x} \in X$. Then the corresponding subsequences of $\{f(x_n)\}$ and $\{z_n\}$ both converge to $f(\hat{x})$. Since f is open and locally one-to-one, and since X is compact, we see easily that we can choose $\epsilon > 0$ so that f maps $N_\epsilon(\hat{x})$ homeomorphically onto an open neighborhood of $f(\hat{x})$, stretching all distances by at least λ . For n sufficiently large, z_n will belong to $f(N_\epsilon(\hat{x}))$, hence $z_n = f(y_n)$ for some uniquely determined $y_n \in N_\epsilon(\hat{x})$. Then $\mathbf{d}(x_n\,,\, y_n) \geq 1/n$ by hypothesis, hence $\mathbf{d}(f(x_n)\,,\, z_n) = \mathbf{d}(f(x_n)\,,\, f(y_n)) \geq \lambda/n$, contradicting the choice of x_n and z_n . □

---

§4F. Problems for the Reader.

Problem 4-a. Chain recurrence. With definitions as in §4B, show that the set $\mathcal{CR}(f)$ of all chain recurrent points is a closed set, and that it contains the non-wandering set $\Omega(f)$. Show also that each chain component is a closed set, and show that the set of all pairs (x, y) with $x \succ y$ forms a closed subset of $X \times X$. If X is compact, then (1) show that this partial ordering does not depend on the choice of metric; and (2) show that $\mathcal{CR}(f^{\circ 2}) = \mathcal{CR}(f)$, or more generally that $\mathcal{CR}(f^{\circ k}) = \mathcal{CR}(f)$. If f has the shadowing property of Problem 3-c, show that $\Omega(f) = \mathcal{CR}(f)$. On the other hand, for the map of Figure 21 on the circle, show that every point is chain recurrent, but only one point is non-wandering.

Problem 4-b. Chain recurrence for a flow or semi-flow. Let X be a metric space, and let $\{f_t : X \to X\}_{t \geq 0}$ be a one parameter semi-group of mappings of X into itself, continuous in both variables. Given $\epsilon > 0$ and $c > 0$, by an (ϵ, c)-chain from x to y will be meant a sequence $x_0, x_1, \ldots, x_n$ of length $n \geq 1$ with $x = x_0$ and $y = x_n$ such that, for each $0 \leq i < n$ there exists a number $t_i \geq c$ so that

$$\mathbf{d}(f_{t_i}(x_i), x_{i+1}) < \epsilon .$$

Define the *chain partial ordering* by saying that $x \succ y$ if and only if there exists some constant $c > 0$ with the following property: for every $\epsilon > 0$ there should exist an (ϵ, c)-chain from x to y. In particular, define x to be *chain recurrent*[3] if and only if $x \succ x$. Show that this relation is transitive, and that the set of (x, y) with $x \succ y$ forms a closed subset of $X \times X$.

Problem 4-c. Many minimal sets for the angle doubling map. The notation $n!$ will be used for n factorial. Any sequence $A = (a_1, a_2, \ldots, a_{n!})$ consisting of $n!$ zeros and ones can also be considered as a sequence $A = (B_1, \ldots, B_n)$ where each $B_j \in \{0,1\}^{(n-1)!}$ is a block consisting of $(n-1)!$ zeros and ones. For each $n \geq 2$ we will inductively choose three of the possible blocks of length $n!$ to be called "admissible", and will choose just one of these to be the "marker" of length $n!$.

> For $n = 2$, the block $(1,1)$ is chosen as the marker, and the blocks $(0,0)$ and $(0,1)$ are also admissible.
>
> For $n \geq 3$, consider blocks of the form $(B_1, \ldots, B_n)$ where each B_j is an admissible $(n-1)!$-block, but only B_1 is the $(n-1)!$-marker block. Choose three of these to be the admissible blocks of length $n!$, and choose the largest in lexicographical order to be the $n!$-marker block.

Thus a sequence of the form $(1, 1, 0, \ldots)$ occurs at the beginning of an admissible $3!$-block and nowhere else. Similarly, for each $n \geq 4$, the beginning of any admissible $n!$-block can be uniquely recognized.

Now let $\mathbf{b} \in \{0,1\}^{\mathbb{N}}$ be the infinite sequence of zeros and ones whose initial $n!$-block is equal to the marker block for each $n \geq 2$, and let

$$M(\mathbf{b}) = \omega(\mathbf{b}, \sigma) \subset \{0,1\}^{\mathbb{N}}$$

be the ω-limit set of $\mathbf{b}$ under the one-sided shift map $\sigma : \{0,1\}^{\mathbb{N}} \to \{0,1\}^{\mathbb{N}}$. (See §1D.) Then $M(\mathbf{b})$ is a minimal set. In fact, for every $n \geq 3$ every sequence $\mathbf{b}' \in M(\mathbf{b})$ contains

[3] The definition given for example in [Guckenheimer and Holmes] is similar, but takes c identically equal to $+1$. It may seem unnatural to give preference to the particular time $t = 1$. However, one usually assumes that X is compact, and in that case their definition of chain recurrence is completely equivalent to the one given above. For a quite different version of the definition, see [Ruelle 1989].

infinitely many copies of the $n!$-marker block. (These copies repeat periodically with period $(n+1)!$.) Hence we can always choose some iterate σ_n of the shift map so that the shifted sequence $\sigma_n(\mathbf{b}')$ starts with this $n!$-marker block. Now as $n \to \infty$ it follows that the elements $\sigma_n(\mathbf{b}') \in \{0,1\}^{\mathbb{N}}$ converge to the limit $\mathbf{b}$. This implies that $M(\mathbf{b})$ is a minimal set. Note that every minimal set constructed in this way is infinite.

There are uncountably many different ways of choosing which blocks are to be called "admissible", and hence uncountably many ways of choosing $\mathbf{b}$. Show that the corresponding uncountably many minimal sets $M(\mathbf{b})$ are pairwise disjoint.

Finally, identifying each element of $\{0,1\}^{\mathbb{N}}$ with the binary expansion of an angle in $\mathbb{R}/\mathbb{Z}$, conclude that the doubling map on $\mathbb{R}/\mathbb{Z}$ also has uncountably many distinct minimal sets. (For sharper and more general statements of this type, see [Boyland].)

Problem 4-d. Measure of meager sets. Given any number $0 < \lambda \le 1$, show that there exists a disjoint collection of open intervals of total length λ contained in the unit interval I so that their union is dense in I. Conclude that the Lebesgue measure of a countable intersection of dense open subsets of I, or of a meager subset of I, can be any number satisfying $0 \le \lambda \le 1$.

Problem 4-e. Circle maps of degree -1. Show that a homeomorphism f of $\mathbb{R}/\mathbb{Z}$ of degree -1 has exactly two fixed points, and that an interval bounded by these two points maps homeomorphically onto itself under $f \circ f$.

Problem 4-f. Forward expansiveness of $f^{\circ k}$. Using uniform continuity, show that if f is forward expansive then every iterate $f^{\circ k}$ is also. (Note: If $f^{\circ k}$ is forward expansive, then it follows trivially that f is also.)

Problem 4-g. Local expansion of $f^{\circ k}$. If f is locally λ-expanding, show that the iterate $f^{\circ k}$ is locally λ^k-expanding for every $k > 0$. Conversely, if $f^{\circ k}$ is locally λ^k-expanding, show that f is locally λ-expanding under the new metric

$$\widehat{\mathbf{d}}(x,y) \;=\; \sum_{i=0}^{k-1} \mathbf{d}(f^{\circ i}(x)\,,\, f^{\circ i}(y))/\lambda^i \,,$$

which is also compatible with the given topology.

Problem 4-h. Differentially expanding maps. (Compare [Shub].) Let M be a compact Riemannian manifold with Riemannian distance function $\mathbf{d}(x,y)$. For a C^1-smooth map $f : M \to M$, show that the following three conditions are equivalent:

(1) There are constants $c > 0$ and $\lambda > 1$ so that

$$\|Df^{\circ n}(v)\| \;\ge\; c\,\lambda^n\,\|v\|$$

for every tangent vector v and every $n \ge 0$.

(2) Some iterate $f^{\circ k}$ is locally expanding.

(3) The map f itself is locally expanding with respect to the distance function associated with the new Riemannian metric

$$\|v\|' \;=\; \sqrt{\|v\|^2 + \|Df(v)\|^2 + \cdots + \|Df^{\circ k-1}(v)\|^2}\;.$$

Problem 4-i. Locally distance increasing implies forward expansive. If $\mathbf{d}(f(x), f(y)) > \mathbf{d}(x,y)$ whenever $\mathbf{d}(x,y) < 2\epsilon$, for a map of a compact metric space, show that any two distinct orbits $x_0 \mapsto x_1 \mapsto \cdots$ and $y_0 \mapsto y_1 \mapsto \cdots$ must be separated by a distance $\mathbf{d}(x_n, y_n) > \epsilon$ for some $n \ge 0$.

Problem 4-j. Non-smooth conjugacies. Given any complex constant a with $|a| < 1$, show that the "Blaschke product"

$$f(z) = z(z-a)/(1-\bar{a}\,z)$$

carries the unit circle $|z| = 1$ onto itself with derivative

$$\frac{d\,\log f(z)}{d\log z} = 1 + \frac{1-a\bar{a}}{\|z-a\|^2} > 1 .$$

Conclude that this map of the unit circle is topologically conjugate to the doubling map on $\mathbb{R}/\mathbb{Z}$. If a is real, show that the multiplier $df(z)/dz$ at the fixed point $z = 1$ is equal to $2/(1-a)$, and conclude that no two of these maps are differentiably conjugate.

Problem 4-k. Expansive homeomorphisms. By definition, a homeomorphism f of a compact metric space is *expansive* if there exists an $\epsilon > 0$ so that for every $x \neq y$ there exists an integer n (which now may be either positive, negative or zero) so that

$$\mathbf{d}(f^{\circ n}(x), f^{\circ n}(y)) > \epsilon .$$

For $k > 1$, show that $f^{\circ k}$ is expansive if and only if f is expansive. Show that the two sided shift (Problem 2-f) is expansive but not forward expansive, and that the map F_1 on the solenoid (§2E) is expansive but not forward expansive. Similarly show that torus automorphisms as descibed in Example 3 of §4D, with $ad - bc = \pm 1$ and $|a + d| > 1 + (ad - bc)$, are expansive but not forward expansive.

Problem 4-ℓ. Expansive, with an attracting fixed point. If X is the countable set consisting of points $a_n = n/\sqrt{1+n^2}$ for $n \in \mathbb{Z}$, together with the two limit points $a_{\pm\infty} = \pm 1$, show that the shift map $a_n \mapsto a_{n+1}$ is expansive (but not forward expansive), with $+1$ as an attracting fixed point.

Note: Here is is essential that X is a non locally connected set. According to [Lewowicz], no expansive homeomorphism of a compact manifold can have a Liapunov stable point. In fact I understand from a private communication that Lewowicz is able to prove the following sharper statement: *Every expansive homeomorphism of a compact Riemannian manifold exhibits sensitive dependence.*

J. Milnor, Stony Brook