Topological Dynamics: A Survey

Ethan Akin
Mathematics Department
The City College
137 Street and Convent Avenue
New York City, NY 10031, USA

September, 2007

Topological Dynamics: Outline of Article

Glossary

Attractor An invariant subset for a dynamical system such that points sufficiently close to the set remain close and approach the set in the limit as time tends to infinity.

Dynamical System A model for the motion of a system through time. The time variable is either discrete, varying over the integers, or continuous, taking real values. Our systems are deterministic, rather than stochastic, so the the future states of the system are functions of the past.

Equilibrium An equilibrium, or a fixed point, is a point which remains at rest for all time.

Invariant Set A subset is invariant if the orbit of each point of the set remains in the set at all times, both positive and negative. The set is + invariant, or forward invariant, if the forward orbit of each such point remains in the set.

Lyapunov Function A continuous, real-valued function on the state space is a Lyapunov function when it is non-decreasing on each orbit as time moves forward.

Orbit The orbit of an initial position is the set of points through which the system moves as time varies positively and negatively through all values. The forward orbit is the subset associated with positive times.

Recurrence A point is recurrent if it is in its own future. Different concepts of recurrence are obtained from different interpretations of this vague description.

Repellor A repellor is an attractor for the reverse system obtained by changing the sign of the time variable.

Transitivity A system is transitive if every point is in the future of every other point. Periodicity, minimality, topological transitivity and chain transitivity are different concepts of transitivity obtained from different interpretations of this vague description.

1 Introduction and History

The many branches of dynamical systems theory are outgrowths of the study of differential equations and their applications to physics, especially celestial mechanics. The classical subject of ordinary differential equations remains active as a topic in analysis, see, e. g. the older texts Coddington and Levinson [28] and Hartman [39] as well as Murdock [51]. One can observe the distinct fields of differentiable dynamics, measurable dynamics (that is, ergodic theory) and topological dynamics all emerging in the work of Poincaré on the Three Body Problem(for a history, see Barrow-Green [23]).

The transition from the differential equations to the dynamical systems viewpoint can be seen in the two parts of the great book Nemitskii and Stepanov [53]. We can illustrate the difference by considering the *initial value problem* in ordinary differential equations:

$$\begin{aligned} \frac{dx}{dt} &= \xi(x) \\ x(0) &= p. \end{aligned} \tag{1.1}$$

Here x is a vector variable in a Euclidean space $X = \mathbb{R}^n$ or in a manifold X, and the initial point p lies in X. The infinitesimal change $\xi(x)$ is thought of as a vector attached to the point x so that ξ is a vector field on X.

The associated *solution path* is the function ϕ such that as time t varies, $x = \phi(t, p)$ moves in X according to the above equation and with $p = \phi(0, p)$ so that p is associated with the initial time $t = 0$. The solution is a curve in the space X along which x moves beginning at the point p. A theorem of differential equations asserts that the function ϕ exists and is unique, given mild smoothness conditions, e. g. Lipschitz conditions, on the function ξ.

Because the equation is *autonomous*, i. e. ξ may vary with x, but is assumed independent of t, the solutions satisfy the following *semigroup identities*, sometimes also called the *Kolmogorov equations*:

$$\phi(t, \phi(s, p)) = \phi(t + s, p). \tag{1.2}$$

Suppose we solve equation (1.1), beginning at p, and after s units of time, we arrive at $q = \phi(s, p)$. If we again solve the equation, beginning now at q, then the identity (1.2) says that we continue to move along the old curve at the same speed. Thus, after t units of time we are where we would have been on the old solution at the same time, $t + s$ units after time 0.

The initial point p is a parameter here. For each solution path it remains constant, the fixed base point of the path. The solution path based at p is also called the *orbit* of p when we want to emphasize the role of the initial point. It follows from the semigroup identities that distinct solution curves, regarded as subsets of X, do not intersect and so X is subdivided, *foliated*, by these curves. Changing p may shift us from one curve to another, but the motion given by the original differential equation is always on one of these curves.

The gestalt switch to the dynamical systems viewpoint occurs when we reverse the emphasis between t and p. Above we thought of p as a fixed parameter and t as the time variable along the solution path. Instead, we now think of the initial point, relabeled x, as our variable and the time value as the parameter.

For each fixed t value we define the time-t map $\phi^t : X \to X$ by $\phi^t(x) = \phi(t, x)$. For each point $x \in X$ we ask whither it has moved in t units of time. The function $\phi : T \times X \to X$ is called the *flow* of the system and the semigroup identities can be rewritten:

$$\phi^t \circ \phi^s \quad = \quad \phi^{t+s} \qquad \text{for all } t, s \in T. \tag{1.3}$$

These simply say that the association $t \mapsto \phi^t$ is a group homomorphism from the additive group T of real numbers to the automorphism group of X. In particular, observe that the time-0 map ϕ^0 is the identity map 1_X.

We originally obtained the flow ϕ by solving a differential equation. This requires differentiable structure on the underlying space which is why we specified that X be a Euclidean space or a manifold. The automorphism group is then the group of *diffeomorphisms* on X.

For the subject of topological dynamics we begin with a flow, a continuous map ϕ subject to the condition (1.3). In modern parlance a flow is just a continuous group action of the group T of additive reals on the topological space X. The automorphism group is the group of *homeomorphisms* on X. If we replace the group of reals by letting T be the group of integers then the action is entirely determined by the generator $f =_{def} \phi^1$, the time 1 homeomorphism with ϕ^n obtained by iterating f n times if n is positive and iterating the inverse f^{-1} $|n|$ times if n is negative.

Above I mentioned the requirement that the vector field ξ be smooth in order that the flow function ϕ be defined. However, I neglected to point out that in the absence of some sort of boundedness condition the solution path might go to infinity in a finite time. In such a case the function ϕ would not be defined on the entire domain $T \times X$ but only on some open subset containing $\{0\} \times X$. The problems related to this issue are handled in the general theory by assuming that the space X is compact. Of course, Euclidean space is only locally compact. In applying the general theory to systems on a noncompact space one usually restricts to some compact invariant subset or else compactifies, i. e. embeds the system in one on a larger, compact space. Already in Nemytskii and Stepanov [53] much attention is devoted to conditions so that a solution path has a compact closure, see also Bahtia and Szego [27].

As topological dynamics has matured the theory has been extended to cover the action of more general topological groups T, usually countable and discrete, or locally compact, or *Polish* (admits a complete, separable metric). This was already emphasized in the first treatise which explicitly concerned topological dynamics, Gottschalk and Hedlund [38].

In differentiable dynamics we return to the case where the space X is a manifold and the flow is smooth. The breadth and depth of the results then obtained make it much more than a subfield of topological dynamics, see, for example, Katok and Hasselblatt [45]. For measurable dynamics we weaken the assumption of continuity to mere measurability but assume that the space carries a measure invariant with respect to the flow, see, for example, Peterson [55] and Rudolph [57]. The measurable and topological theories are especially closely linked with a number of parallel results, see Glasner [36] as well as Alpern and Prasad [12]. The reader should also take note of Oxtoby [54], a beautiful little book which explicitly describes this parallelism using a great variety of applications.

The current relationship between topological dynamics and dynamical systems theory in general is best understood by analogy with that between point-set, or general, topology and analysis.

General topology proper, even excluding algebraic topology and homotopy theory, is a large specialty with a rich history and considerable current research (for some representative surveys see the Russian Encyclopedia volumes [13],[14] and [15]). But much of this work is little known to nonspecialists. On the other hand, the fundamentals of point-set topology are part of the foundation upon which modern analysis is built. Compactness was a rather new idea when it was used by Jesse Douglas in his solution of the Plateau Problem, Douglas [32] (see Almgren [11]). Nowadays continuity, compactness and connectedness are in the vocabulary of every analyst. In addition, there recur unexpected applications of hitherto specialized topics. For example, indecomposable continua, examined by Bing and his students, see Bing [25], are now widely recognized and used in the study of strange attractors, see Brown [26].

Similarly, the area of topological dynamics proper is large and some of the more technical results have found application. We will touch on some of these in the end, but for the most part this article will concentrate on those basic aspects which provide a foundation for dynamical systems theory in general. We will focus on chain recurrence and the theory of attractors, following the exposition of Akin [1] and Akin, Hurley and Kennedy [8]).

To describe what we want to look for, let us begin with the simplest qualitative situation: the differential equation model (1.1) where the vector field ξ is the gradient of some smooth real-valued potential function U on X. Think of X as the Euclidean plane and the graph of U as a surface in space over X. The motion in X can be visualized on the surface above. On the surface it is always upward, perpendicular to the contour curves of constant height. For simplicity we will assume that U has isolated critical points. These critical points: local maxima, minima and saddles, are *equilibria* for the system, points at which the the gradient field ξ vanishes. We observe two kinds of behavior. The orbit of a critical point is constant, resting at equilibrium. The other kind exhibits what engineers call *transient* behavior. A non-equilibrium solution path moves asymptotically toward a critical point (or towards infinity). As it approaches its limit, the motion slows, becoming imperceptible, indistinguishable from rest at the limit point. The set of points whose orbits tend to a particular critical point e is called the *stable set* for e.

Each local maximum e is an *attractor* or *sink*. The stable set for e is an open set containing e which is called the *domain of attraction* for e. Such a state is called *asymptotically stable* illustrated by the rest state of a cone on its base.

The local minima are *repellors* or *sources* which are attractors for the system with time reversed. Solution paths near a repellor move away from it. The stable set for a repellor e consists of e alone. Consider a cone balanced on its point.

A saddle point between two local maxima is like the highest point of a pass between two mountains. Separating the domains of attraction for the two peaks are solution paths which have limit the saddle point equilibrium.

There is a kind of knife, used for cutting bread dough, which has a semicircular blade. The saddle point equilibrium is a state like this knife balanced on the midpoint of its blade. A slight perturbation will cause it to fall down on one side or the other. But if you start the knife balanced elsewhere along its blade there remains the possibility - not achievable in practice- of its rolling back along the blade toward balance at the midpoint equilibrium. Notice that these are first-order systems with no momentum. Imagine everything going on in thick, clear molasses. As this model illustrates, it is usually true that the stable set of a saddle point is a lower dimensional set in X and the union of the domains of attraction of the local maxima is a dense open subset of X.

There do exist examples where the stable set of a saddle has nonempty interior. This is a pathology which we will hope to exclude by imposing various conditions. For example, the potential function U is called a *Morse function* when all of its critical points are nondegenerate. That is, the *Hessian* matrix of second partials is nonsingular at each critical point. For the gradient system of a Morse function each equilibrium is of a type called *hyperbolic*. For the saddle points of such a system the stable sets are manifolds of lower dimension.

From the cone example, we omitted what physicists call *neutral stability*, the cone resting on its side. From our point of view this is another sort of pathology: an infinite, connected set of equilibria. Each of these equilibria is *stable* but not asymptotically stable. If we perturb the cone by lifting its point and turning it a bit, then it drops back toward an equilibrium near to but not necessarily identical with the original state (Remember, no momentum).

We obtain a similar classification into sinks, saddles, etc. and complementary transient behavior when we remove the assumption that the system comes from the gradient of U and retain only the condition that U increases along nonequilibrium solution paths. The function U is then called a *strict Lyapunov function* for the system. Instead of the steepest path ascent of a mountain goat, we may observe the spiralling upward of a car on a mountain road.

However, these gradient-like systems are too simple to represent a typical dynamical system. Lacking is the general behavior complementary to transience, namely *recurrence.* A point is recurrent - in some sense - if it is “in its own future” - in the appropriate sense. Beyond equilibrium the simplest kind of recurrence occurs on a periodic orbit. A periodic orbit returns infinitely often to each point on the orbit. As we will see, there are increasingly broad concepts of recurrence obtained by extending the notion of the “future” of a point. Clearly, a real-valued function cannot be strictly increasing along a periodic orbit. When we consider Lyapunov functions in general we will see that they remain constant along the orbit of each recurrent point.

Nonetheless, the picture we are looking for can be related to the gradient landscape by replacing the critical points by blobs of various sizes. Each blob is a closed, invariant set of a special type. A subset A is an *invariant set* for the system when A contains the entire orbit of each of its points. The special condition on each blob A which replaces an equilibrium is a kind of *transitivity.* This means that if p and q are points of A then each point is in the “future” of the other and so, in particular, each point is recurrent in the appropriate sense. Here again it remains to provide a meaning - or actually several different meanings - for this vague notion of “future”.

In this more general situation the transient orbits need not converge to a point. Instead, each accumulates on a closed subset of one of these blobs. If there are only finitely many of the blobs then there is a classification of them as attractor, repellor, or saddle analogous to the description for equilibria in the gradient system.

Within each blob the motion may be quite complicated. It is in attempting to describe such motions that the concept of *chaos* arises.

For some applied fields this sort of thinking is relatively new. When I learned population genetics - admittedly that was over thirty years ago - most of the analysis consisted of identifying and classifying the equilibria, tricky enough in several variables. This is perfectly appropriate for gradient-like systems and is a good first step in any case. However, it has become apparent that more complicated recurrence may occur and so requires attention.

While most applications use differential equations and the associated flows, it is more convenient to develop the discrete time theory and then to derive from it the results for flows. In what follows we will describe the results for a *cascade*, a homeomorphism f on a compact metric space X with the dynamics introduced by iteration. Focusing on this case, we will not discuss further real flows or noninvertible functions, and we will omit as well the extensions to noncompact state spaces and to compact, non-metrizable spaces.

2 Dynamic Relations, Invariant Sets and Lyapunov Functions

It is convenient to assume that our state spaces are nonempty and metrizable. It is essential to assume that they are compact. Recall that if A is a subset of a compact metric space X then A is closed if and only if it is compact. Also, the continuous image of a compact set is compact and so for continuous maps between compact metric spaces the image of a closed set is closed. Perhaps less familiar are the following important results:

Proposition 2.1 *Let X be a compact metric space and $\{A_n\}$ be a decreasing sequence of closed subsets of X with intersection A.*

(a) If U is an open subset of X with $A \subseteq U$ then for sufficiently large n, $A_n \subseteq U$.

(b) If A_n is nonempty for every n, then the intersection A is nonempty.

(c) If $h : X \to Y$ is a continuous map with Y a metric space then

$$\bigcap_n h(A_n) \quad = \quad h(A). \tag{2.1}$$

Proof: (a): We are assuming $A_{n+1} \subseteq A_n$ for all n and $A = \bigcap_n A_n$. The complementary open sets $V_n = X \setminus A_n$ are increasing and, together with U, they cover X. By compactness, $\{U, V_1, ..., V_N\}$ covers X for some N and so $\{U, V_N\}$ suffice to cover X. Hence, A_N is a subset of U as is A_n for any $n \geq N$.

(b): If A is empty then $U = \emptyset$ is an open set containing A and so by (a), A_n is empty for sufficiently large n.

(c): Since $A \subseteq A_n$ for all n, it is clear that $h(A)$ is contained in $\bigcap_n h(A_n)$. On the other hand, if y is a point of the latter intersection then $\{h^{-1}(y) \cap A_n\}$ is a decreasing sequence of nonempty compact sets. By (b) the intersection $h^{-1}(y) \cap A$ is nonempty. QED

If $\{B_n\}$ is any sequence of closed subsets of X then we define the *Lim sup* :

$$Limsup_n\ B_n \quad =_{def} \quad \bigcap_n \overline{\bigcup_{k \geq n} B_k} \tag{2.2}$$

where $\overline{Q}$ denotes the closure of Q. It follows from (b) above that the Lim sup of a sequence of nonempty sets is nonempty. It is easy to check that

$$\overline{\bigcup_n B_n} \quad = \quad (\bigcup_n B_n) \cup (Limsup_n \, B_n). \tag{2.3}$$

We want to study a homeomorphism on a space. By compactness this is just a bijective (= one-to-one and onto) continuous function from the space to itself. It will be convenient to use the more general language of relations.

A function $f : X \to Y$ is usually described as a rule associating to every point x in X a unique point $y = f(x)$ in Y. In set theory the function f is defined to be the set of ordered pairs $\{(x, f(x)) : x \in X\}$. Thus, the function f is a subset of the product $X \times Y$. It is what other people call the graph of the function. We will use this language so that, for example, the identity map 1_X on X is the diagonal subset $\{(x,x) : x \in X\}$. The notation is extended by defining a *relation from* X *to* Y, written $F : X \to Y$, to be an arbitrary subset of $X \times Y$. Then $F(x) = \{y : (x,y) \in F\}$. Thus, a relation is a function exactly when the set $F(x)$ contains a single point for every $x \in X$. In the function case, we will use the symbol $F(x)$ for both the set and the single point it contains, the latter being the usual meaning of $F(x)$.

As they are arbitrary subsets of $X \times Y$ we can perform set operations like union, intersection, closure and interior on relations. In addition, for $F : X \to Y$ we define the *inverse* $F^{-1} : Y \to X$ by

$$F^{-1} \quad =_{def} \quad \{(y,x) : (x,y) \in F\}. \tag{2.4}$$

If $A \subseteq X$ then its *image* is

$$\begin{aligned} F(A) \quad &=_{def} \quad \{y : (x,y) \in F \text{ for some } \ x \in A\} \\ &= \quad \bigcup_{x \in A} F(x) \quad = \quad \pi_2((A \times Y) \cap F), \end{aligned} \tag{2.5}$$

where $\pi_2 : X \times Y \to Y$ is the projection to the second coordinate. It follows that if $B \subseteq Y$ then

$$\begin{aligned} F^{-1}(B) \quad &= \quad \{x : F(x) \cap B \; \neq \; \emptyset\}, \\ X \setminus F^{-1}(Y \setminus B) \quad &= \quad \{x : F(x) \subseteq B\}. \end{aligned} \tag{2.6}$$

These two sets are usually different but they agree when F is a function.

If $G : Y \to Z$ is another relation then the *composition* $G \circ F : X \to Z$ is the relation given by

$$\begin{aligned} G \circ F \quad &=_{def} \quad \{(x,z) : \text{ there exists } y \in Y \\ &\text{such that } (x,y) \in F \text{ and } (y,z) \in G\} \\ &= \quad \pi_{13}((X \times G) \cap (F \times Z)), \end{aligned} \tag{2.7}$$

where $\pi_{13} : X \times Y \times Z \to X \times Z$ is the projection map. This generalizes composition of functions and, as with functions, composition is associative. Clearly, $(G \circ F)^{-1} = F^{-1} \circ G^{-1}$.

We call F a *closed relation* when it is a closed subset of $X \times Y$. Clearly, the inverse of a closed relation is closed and by compactness, the composition of closed relations is closed. If A is a closed subset of X and F is a closed relation then the image $F(A)$ is a closed subset of Y. It follows from (2.6) that if B is an open subset of Y then $\{x \in X : F(x) \subseteq B\}$ is an open subset of X. Thus, for relations being closed is analogous to being continuous for functions. In fact, a function is continuous if and only if, regarded as a relation, it is closed. This is another application of compactness.

If $Y = X$, so that $F : X \to X$, then we call F a *relation on* X. For a positive integer n we define F^n to be the n-fold composition of F with $F^0 =_{def} 1_X$ and $F^{-n} =_{def} (F^{-1})^n = (F^n)^{-1}$. This is well-defined because composition is associative. Clearly, $F^m \circ F^n = F^{m+n}$ when m and n have the same sign, i.e. when $mn \geq 0$. On the other hand, the equations $F \circ F^{-1} = F^{-1} \circ F = 1_X = F^0$ all hold if and only if the relation F is a bijective function.

The utility of this relation-speak, once one gets used to it, is that it allows us to extend to this more general situation our intuitions about a function as a way of moving from input here to output there. For example, if $\epsilon \geq 0$ then we can use the metric d on X to define the relations on X

$$\begin{aligned} V_\epsilon \quad &=_{def} \{(x,y) : d(x,y) < \epsilon\}, \\ \bar{V}_\epsilon \quad &=_{def} \{(x,y) : d(x,y) \leq \epsilon\}. \end{aligned} \tag{2.8}$$

Thus, $V_\epsilon(x)$ and $\bar{V}_\epsilon(x)$ are the open ball and the closed ball centered at x with radius ϵ. We can think of these relations as ways of moving from a point x to a nearby point.

Each $\bar{V}_\epsilon$ is a closed, symmetric and reflexive relation. The triangle inequality is equivalent to the inclusion $\bar{V}_\epsilon \circ \bar{V}_\delta \subseteq \bar{V}_{\epsilon+\delta}$.

In general, a relation F on X is *reflexive* if $1_X \subseteq F$, *symmetric* if $F^{-1} = F$ and *transitive* if $F \circ F \subseteq F$.

Now we apply this relation notation to a homeomorphism f on X.

For $x \in X$ the *orbit sequence* of x is the bi-infinite sequence $\{..., f^{-2}(x), f^{-1}(x), x, f(x), f^2(x), ...\}$. This is just the discrete time analogue of the solution path discussed in Section 1. We are thinking of it as a sequence and so as a function from the set $\mathbb{Z}$ of discrete times to the state space X with parameter the initial point x.

Now we define the *orbit relation*

$$\mathcal{O}f \quad =_{def} \quad \bigcup_{n=1}^{\infty} f^n. \tag{2.9}$$

Thus, $\mathcal{O}f(x) = \{f(x), f^2(x), ...\}$ is a set, not a sequence, consisting of the states which follow the initial point x in time. Notice that for reasons which will be clear when we consider the cyclic sets below, we begin the union with $n = 1$ rather than $n = 0$ and so the initial point x itself need not be included in $\mathcal{O}f(x)$.

If with think of the point $f(x)$ as the immediate temporal successor of x then $\mathcal{O}f(x)$ is the set of points which occur on the orbit of x at some positive time. This is the first -and simplest- interpretation of the "future" of x with respect to the dynamical system obtained by iterating f.

It is convenient to extend this notion by including the limit points of the positive orbit sequence. For $x \in X$ define

$$\begin{aligned} \omega f(x) \quad &= \quad Limsup_n \ \{f^n(x)\}, \\ \mathcal{R}f(x) \quad &=_{def} \quad \overline{\mathcal{O}f(x)} \quad = \quad \mathcal{O}f(x) \cup \omega f(x). \end{aligned} \tag{2.10}$$

Thus, from f we have defined the orbit relation $\mathcal{O}f$ and the *orbit closure relation* $\mathcal{R}f$ with $\mathcal{R}f = \mathcal{O}f \cup \omega f$.

While $\mathcal{R}f(x)$ is closed for each x, the relation $\mathcal{R}f$ itself is usually not closed. As was mentioned above, among relations the closed relations are the analogues of continuous functions. We obtain closed relations by defining

$$\begin{aligned} \Omega f \quad &= \quad Limsup_n \ f^n, \\ \mathcal{N}f(x) \quad &=_{def} \quad \overline{\mathcal{O}f} \quad = \quad \mathcal{O}f \cup \Omega f. \end{aligned} \tag{2.11}$$

Here we are taking the closure in $X \times X$ and so we obtain closed relations. The relation $\mathcal{N}f$, defined by Auslander et al [18], [19] and [20] (see also Ura [64],[65]), is called the *prolongation* of f. $\mathcal{N}f(x)$ is our next, broader, notion of the "future" of x.

To compare all these suppose that $x, y \in X$. Then $y \in \mathcal{O}f(x)$ when there exists a positive integer n such that $y = f^n(x)$ while $y \in \omega f(x)$ if there is a sequence of positive integers $n_i \to \infty$ such that $f^{n_i}(x) \to y$. On the other hand, $y \in \Omega f(x)$ iff there are sequences $x_i \to x$ and $n_i \to \infty$ such that $f^{n_i}(x_i) \to y$. Thus, $y \in \mathcal{R}f(x)$ if for every $\epsilon > 0$ we can run along the orbit of x and at some positive time make a smaller than ϵ jump to y. Similarly, $y \in \mathcal{N}f(x)$ if for every $\epsilon > 0$ we can make an initial ϵ small jump to a point x_1, run along the orbit of x_1 and at some positive time make an ϵ small jump to y. Thus,

$$\mathcal{R}f \quad = \quad \bigcap_{\epsilon>0} V_\epsilon \circ (\mathcal{O}f) \qquad \text{and} \qquad \mathcal{N}f \quad = \quad \bigcap_{\epsilon>0} V_\epsilon \circ (\mathcal{O}f) \circ V_\epsilon. \tag{2.12}$$

The relations $\mathcal{O}f$ and $\mathcal{R}f$ are transitive but usually not closed. In general, when we pass to the closure, obtaining $\mathcal{N}f$, we lose transitivity. When we consider Lyapunov functions we will see why it is natural to want both of these properties, closure and transitivity.

The intersection of any collection of closed, transitive relations is a closed, transitive relation. Notice that $X \times X$ is such a relation. Thus, we obtain $\mathcal{G}f$, the smallest closed, transitive relation which contains f by intersecting:

$$\mathcal{G}f \quad =_{def} \quad \bigcap \{Q \subseteq X \times X : \overline{Q} \ = \ Q \text{ and } \ f, \ Q \circ Q \subseteq Q\}. \tag{2.13}$$

There is an alternative procedure, due to Conley, see [29], which constructs a closed, transitive relation, generally larger than $\mathcal{G}f$, in a simple and direct fashion.

A *chain* or 0-chain is a finite or infinite sequence $\{x_n\}$ such that $x_{n+1} = f(x_n)$ along the way, i. e. a piece of the orbit sequence. Given $\epsilon \geq 0$ an ϵ-chain is a finite or infinite sequence $\{x_n\}$ such that each x_{n+1} at most ϵ distance away from the point $f(x_n)$, i. e. $x_{n+1} \in \bar{V}_\epsilon(f(x_n))$. If the chain has at least two terms, but only finitely many, then the first and last terms are called the beginning and the end of the chain. The number of terms minus 1 is then called the *length* of the chain. We say that x *chains to* y, written $y \in \mathcal{C}f(x)$ if for every $\epsilon > 0$ there is an ϵ chain which begins at x and ends at y. That is,

$$\mathcal{C}f \quad =_{def} \quad \bigcap_{\epsilon>0} \mathcal{O}(\bar{V}_\epsilon \circ f). \tag{2.14}$$

Compare this with (2.12). If $y \in \mathcal{R}f(x)$ then we can get to y by moving along the orbit of x and then taking an arbitrarily small jump at the end. If

$y \in \mathcal{N}f(x)$ then we are allowed a small jump at the beginning as well as the end. Finally, if $y \in \mathcal{C}f(x)$ then we are allowed a small jump at each iterative step.

The chain relation is of great importance for applications. Suppose we are computing the orbit of a point on a computer. At each step there is usually some round-off error. Thus, what we take to be an orbit sequence is in reality an ϵ chain for some positive ϵ. It follows that, in general, what we can expect to compute directly about f is only that level of information which is contained in $\mathcal{C}f$.

As the intersection of transitive relations, $\mathcal{C}f$ is transitive. It is not hard to show directly that the chain relation $\mathcal{C}f$ is also closed. Since $x_1 = x,\ x_2 = f(x)$ is a 0-chain beginning at x and ending at $f(x)$, we have $f \subseteq \mathcal{C}f$. It follows that $\mathcal{G}f \subseteq \mathcal{C}f$.

This inclusion may be strict. The identity map $f = 1_X$ is already a closed equivalence relation. Hence, $\mathcal{G}1_X = 1_X$. On the other hand, for any $\epsilon > 0$ the relation $\mathcal{O}V_\epsilon$ is an open equivalence relation, and so each equivalence class is clopen (= closed and open). If X is connected then the entire space is a single equivalence class. Since this is true for every positive ϵ we have

$$X \text{ connected} \qquad \Longrightarrow \qquad \mathcal{C}1_X \quad = \quad X \times X. \tag{2.15}$$

Since $\mathcal{C}f$ is transitive, the composites $\{(\mathcal{C}f)^n\}$ form a decreasing sequence of closed transitive relations. We denote by $\Omega\mathcal{C}f$ the intersection of this sequence. That is,

$$\Omega\mathcal{C}f \quad =_{def} \quad \bigcap_{n=0}^{\infty} (\mathcal{C}f)^n. \tag{2.16}$$

One can show that $y \in \Omega\mathcal{C}f(x)$ if and only if for every $\epsilon > 0$ and positive integer N there is an ϵ chain of length greater than N which begins at x and ends at y. In addition, the following identity holds (compare (2.10) and (2.11)):

$$\mathcal{C}f \quad = \quad \mathcal{O}f \cup \Omega\mathcal{C}f. \tag{2.17}$$

Thus, built upon f we have a tower of relations:

$$f \quad \subseteq \quad \mathcal{O}f \quad \subseteq \quad \mathcal{R}f \quad \subseteq \quad \mathcal{N}f \quad \subseteq \quad \mathcal{G}f \quad \subseteq \quad \mathcal{C}f. \tag{2.18}$$

These are the relations which capture the successively broader notions of the "future" of an input x.

A useful identity which holds for $\mathcal{A} = \mathcal{O}, \mathcal{R}, \mathcal{N}, \mathcal{G}$ and $\mathcal{C}$ is

$$\mathcal{A}f \quad = \quad f \cup (\mathcal{A}f) \circ f \quad = \quad f \cup f \circ (\mathcal{A}f). \tag{2.19}$$

These are easy to check directly for all but $\mathcal{G}$. For that one, observe that $f \cup (\mathcal{G}f) \circ f$ and the other composite are closed and transitive and so each contains $\mathcal{G}f$.

It is also easy to show that for $\mathcal{A} = \mathcal{O}, \mathcal{N}, \mathcal{G}$ and $\mathcal{C}$: $\mathcal{A}(f^{-1}) = (\mathcal{A}f)^{-1}$ and so we can omit the parentheses in these cases. The analogue for $\mathcal{R}$ is usually false and we define

$$\alpha f \quad =_{def} \quad \omega(f^{-1}). \tag{2.20}$$

Thus, $\alpha f(x)$ is the set of limit points of the negative time orbit sequence $\{x, f^{-1}(x), f^{-2}(x), ...\}$ and it is usually not true that αf equals $(\omega f)^{-1}$.

For $\Theta = \alpha, \omega, \Omega$ and $\Omega\mathcal{C}$ it is true that

$$\Theta f \quad = \quad f \circ \Theta f \quad = \quad f^{-1} \circ \Theta f \quad = \quad \Theta f \circ f \quad = \quad \Theta f \circ f^{-1}. \tag{2.21}$$

Now we are ready to consider the variety of recurrence concepts. Recall that a point x is recurrent - in some sense - if it lies in its own "future". Thus, for any relation F on X we define the *cyclic set*

$$|F| \quad =_{def} \quad \{x : (x, x) \in F\} \tag{2.22}$$

Clearly, if F is a closed relation then $|F|$ is a closed subset of X.

A point x lies in $|f|$ when $x = f(x)$ and so $|f|$ is the set of *fixed points* for f, while $x \in |\mathcal{O}f|$ when $x = f^n(x)$ for some positive integer n and so $|\mathcal{O}f|$ is the set of *periodic points*. It is easy to check that every periodic point is contained in $|\omega f|$ and so $|\omega f| = |\mathcal{R}f|$. These are called *recurrent points* or sometimes the *positive recurrent points* to distinguish them from $|\alpha f|$, the set of *negative recurrent points*. Similarly, we have $|\Omega f| = |\mathcal{N}f|$, called the set of *nonwandering points*. The points of $|\mathcal{G}f|$ are called *generalized nonwandering* and those of $|\mathcal{C}f| = |\Omega\mathcal{C}f|$ are called *chain recurrent*. The set of periodic points and the sets of recurrent points need not be closed. The rest, associated with closed relations, are closed subsets.

For an illustration of these ideas, observe that for $x, y, z \in X$

$$y, z \ \in \ \omega f(x) \qquad \Longrightarrow \qquad z \ \in \ \Omega f(y). \tag{2.23}$$

Just hop from y to a nearby point on the orbit of x, moving arbitrarily far along the orbit, you repeatedly arrive nearby z and then can hop to it. In particular, with $y = z$ we see that every point y of $\omega f(x)$ is non-wandering. However, the points of $\omega f(x)$ need not be recurrent. That is, while $y \in \Omega f(y)$ for all $y \in \omega f(x)$ it need not be true that $y \in \omega f(y)$. In particular, $\mathcal{R}f(y)$ can be a proper subset of $\mathcal{N}f(y)$.

On the other hand, it is true that for most points x in X the orbit closure $\mathcal{R}f(x)$ is equal to the prolongation set $\mathcal{N}f(x)$. Recall that a subset of a complete metric space is called *residual* when it is the countable intersection of dense, open subsets. By the Baire Category Theorem a residual subset is dense.

Theorem 2.2 *If f is a homeomorphism on X then*
$\{x \in X : \omega f(x) = \Omega f(x)\} \ = \ \{x \in X : \mathcal{R}f(x) = \mathcal{N}f(x)\}$
is a residual subset of X.

In particular, if every point is nonwandering, i. e. $x \in \Omega f(x)$ for all x, then the set of recurrent points is residual in X. However, if the set of nonwandering points is a proper subset of X, it need not be true that most of these points are recurrent. The closure of $|\omega f|$ can be a proper subset of the closed set $|\Omega f|$.

From recurrence we turn to the notion of invariance.

Let F be a relation on X and A be a closed subset of X. We call A *+ invariant* for F if $F(A) \subseteq A$ and *invariant* for F if $F(A) = A$. For a homeomorphism f on X, the set A is invariant for f if and only if it is + invariant for f and for f^{-1}. For example, (2.19) implies that for $\mathcal{A} = \mathcal{O}, \mathcal{R}, \mathcal{N}, \mathcal{G}$ and $\mathcal{C}$ each of the sets $\mathcal{A}f(x)$ is + invariant for f and (2.21) implies that for $\Theta = \alpha, \omega, \Omega$ and $\Omega\mathcal{C}$ each of the sets $\Theta f(x)$ is invariant for f. Finally, for $\mathcal{A} = \mathcal{O}, \mathcal{R}, \mathcal{N}, \mathcal{G}$ and $\mathcal{C}$ each of the cyclic sets $|\mathcal{A}f|$ is invariant for f.

If A is a nonempty, closed invariant subset for a homeomorphism f then the restriction $f|A$ is a homeomorphism on A and we call this dynamical system the *subsystem* determined by A. In general, if F is a relation on X and A is any subset of X then we call the relation $F \cap (A \times A)$ on A the *restriction* of F to A.

For a homeomorphism f the families of + invariant subsets and of invariant subsets are each closed under the operations of closure and interior and under arbitrary unions and intersections. If A is + invariant then the sequence $\{f^n(A)\}$ is decreasing and the intersection is f invariant. Furthermore, this intersection contains every other f invariant subset of A and so is the maximum invariant subset of A.

If A is $+$ invariant for f then it is $+$ invariant for $\mathcal{O}f$. If, in addition, A is closed then it is $+$ invariant for $\mathcal{R}f$. However, $+$ invariance with respect to the later relations in the tower (2.18) are successively stronger conditions, and the relations of (2.18) provide convenient tools for studying these conditions.

We call a closed $+$ invariant subset A a *stable* subset, or a *Lyapunov stable* subset, if it has a neighborhood basis of $+$ invariant neighborhoods. That is, if G is open and $A \subseteq G$ then there exists a $+$ invariant open set U such that $A \subseteq U \subseteq G$.

Theorem 2.3 *A closed subset A is $+$ invariant for $\mathcal{N}f$ if and only if it is a stable $+$ invariant set for f.*

Proof: If G is an open set which contains A and $\mathcal{N}f(A) \subseteq A$ then $U = \{x : \mathcal{N}f(x) \subseteq G\}$ is an open set which contains A and which is $+$ invariant by (2.19). The reverse implication is easy to check directly. QED

Invariance with respect to $\mathcal{G}f$ is characterized by using Lyapunov functions, which generalize the strict Lyapunov functions described Section 1.

For a closed relation F on X, a Lyapunov function L for F is a continuous, real-valued function on X such that

$$(x, y) \in F \qquad \Longrightarrow \qquad L(x) \leq L(y), \tag{2.24}$$

(some authors, e. g. Lyapunov, use the reverse inequality).

For any continuous, real-valued function L on X the set $\{(x, y) : L(x) \leq L(y)\}$ is a closed, transitive relation on X. To say that L is a Lyapunov function for F is exactly to say that this relation contains F. It follows that if L is a Lyapunov function for a homeomorphism f then it is automatically a Lyapunov function for the closed, transitive relation $\mathcal{G}f$. That L be a Lyapunov function for $\mathcal{C}f$ is usually a stronger condition. For example, any continuous, real-valued function is a Lyapunov function for 1_X, but by (2.15) if X is connected then constant functions are the only Lyapunov functions for $\mathcal{C}1_X$.

The following result is a dynamic analogue of Urysohn's Lemma in general topology and it has a similar proof.

Theorem 2.4 *A closed subset A is + invariant for $\mathcal{G}f$ if and only if there exists a Lyapunov function $L : X \to [0,1]$ for f such that $A = L^{-1}(1)$.*

A Lyapunov function for a homeomorphism f is non-decreasing on each orbit sequence $x, f(x), f^2(x), ...$ Suppose that x is a periodic point, i.e. $x \in |\mathcal{O}f|$. Then $x = f^n(x)$ for some positive integer n, and it follows that L must be constant on the orbit of x.

Now suppose, more generally, that x is a generalized recurrent point, i.e. $x \in |\mathcal{G}f|$. By (2.19) $(f(x), x) \in \mathcal{G}f$ and so $L(f(x)) = L(x)$ whenever L is a Lyapunov function for f (and hence for $\mathcal{G}f$). Recall that the set $|\mathcal{G}f|$ of generalized recurrent points is f invariant. It follows that $f^n(x) \in |\mathcal{G}f|$ for every integer n and so $L(f^{n+1}(x)) = L(f^n(x))$ for all n. Thus, a Lyapunov function for a homeomorphism f is constant on the orbit of each generalized recurrent point x.

Similarly, if x is a chain recurrent point, i.e. $x \in |\mathcal{C}f|$, then $L(f(x)) = L(x)$ whenever L is a Lyapunov function for $\mathcal{C}f$ and so a Lyapunov function for $\mathcal{C}f$ is constant on the orbit of each chain recurrent point.

Of fundamental importance is the observation that one can construct Lyapunov functions for f (and for $\mathcal{C}f$) which are increasing on all orbit sequences which are not generalized recurrent (respectively, chain recurrent).

Theorem 2.5 *For a homeomorphism f on a compact metric space X there exist continuous functions $L_1, L_2 : X \to [0,1]$ such that L_1 is a Lyapunov function for f, and hence for $\mathcal{G}f$, and L_2 is a Lyapunov function for $\mathcal{C}f$ and, in addition,*

$$\begin{array}{lcrcl} x \in |\mathcal{G}f| & \iff & L_1(x) & = & L_1(f(x)), \\ x \in |\mathcal{C}f| & \iff & L_2(x) & = & L_2(f(x)). \end{array} \tag{2.25}$$

Lyapunov functions which satisfy the conditions of (2.25) are called *complete Lyapunov functions* for f and for $\mathcal{C}f$, respectively.

If L is a Lyapunov function for f then we define the set of *critical points* for L

$$|L| \quad =_{def} \quad \{x \in X : \quad L(x) \ = \ L(f(x))\}. \tag{2.26}$$

This language is a bit abusive because here criticality describes a relationship between L and f. It does not depend only upon L. However, we adopt this language to compare the general situation with the simpler strict Lyapunov function case in Section 1. Similarly, we call $L(|L|) \subseteq \mathbb{R}$ the set of *critical values* for L. The complementary points of X and $\mathbb{R}$ respectively are called *regular points* and *regular values* for L. Thus, the generalized recurrent points are always critical points for a Lyapunov function L and for a complete Lyapunov function these are the only critical points.

For invariance with respect to $\mathcal{C}f$ we turn to the study of attractors.

3 Attractors and Chain Recurrence

For a homeomorphism f on X we say that a closed set U is *inward* if $f(U)$ is contained in U°, the interior of U. By compactness this implies that $\bar{V}_\epsilon \circ f(U) \subseteq U$ for some $\epsilon > 0$. That is, U is + invariant for the relation $\bar{V}_\epsilon \circ f$. Hence, any ϵ chain for f which begins in U remains in U. It follows that an inward set for f is $\mathcal{C}f$ + invariant.

For example, assume that L is a Lyapunov function for f. Then for any $s \in \mathbb{R}$ the closed set $U_s = \{x : L(x) \geq s\}$ is + invariant for f. Suppose now that s is a regular value for L. This means that for all x such that $L(x) = s$ we have $L(f(x)) > L(x) = s$. On the other hand, for the remaining points x of U_s we have $L(f(x)) \geq L(x) > s$. Thus, $f(U_s)$ is contained in the open set $\{x : L(x) > s\} \subseteq U_s$ and so U_s is inward. It easily follows that L is a Lyapunov function for $\mathcal{C}f$ if the set of critical values is nowhere dense.

If U is inward for f then we define $A = \bigcap_{n=0}^{\infty}\{f^n(U)\}$ to be the *attractor* associated with U. Since an inward set U is + invariant for f, the sequence $\{f^n(U)\}$ is decreasing. In fact, for U inward and $n, m \in \mathbb{Z}$ we have

$$n > m \qquad \Longrightarrow \qquad f^n(U) \subseteq f^m(U)^\circ. \tag{3.1}$$

The associated attractor is the maximum f invariant subset of U and $\{f^n(U) : n \in \mathbb{Z}\}$ is a sequence of inward neighborhoods of A, forming a neighborhood basis for the set A. That is, if G is any open which contains A then by Proposition 1.1a, $f^n(U) \subseteq G$ for sufficiently large n.

For example, the entire space X is an inward set and is its own associated attractor. In general, a set A is inward and equal to its own attractor if and only if A is a clopen, f invariant set.

The power of the attractor idea comes from the equivalence of a number of descriptions of different apparent strength. We use the "weak" ones to test for an attractor and then apply the "strong" conditions. The following collects these alternative descriptions.

Theorem 3.1 *Let f be a homeomorphism on X and A be a closed f invariant subset of X. The following conditions are equivalent.*

(i) A is an attractor. That is, there exists an inward set U such that $\bigcap_{n=0}^{\infty} f^n(U) = A$.

(ii) There exists a neighborhood G of A such that $\bigcap_{n=0}^{\infty} f^n(G) \subseteq A$.

(iii) A is $\mathcal{N}f$ + invariant and the set $\{x : \omega f(x) \subseteq A\}$ is a neighborhood of A.

(iv) The set $\{x : \Omega f(x) \subseteq A\}$ is a neighborhood of A.

(v) The set $\{x : \Omega \mathcal{C} f(x) \subseteq A\}$ is a neighborhood of A.

(vi) A is $\mathcal{G}f$ + invariant and the set $A \cap |\mathcal{G}f|$ is clopen in the relative topology of the closed set $|\mathcal{G}f|$.

(vii) A is $\mathcal{C}f$ + invariant and the set $A \cap |\mathcal{C}f|$ is clopen in the relative topology of the closed set $|\mathcal{C}f|$.

Applying Theorem 2.3 to condition (iii) of Theorem 3.1 we see that if A is a stable set for f and, in addition, the orbit of every point in some neighborhood of x tends asymptotically toward A then A is an attractor. The latter condition alone does not suffice, although the strengthening in (iv) is sufficient. For example, the homeomorphism of $[0, 1]$ defined by $t \mapsto t^2$ has $\{0\}$ as an attractor. If we identify the two fixed points $0, 1 \in [0, 1]$ by mapping t to $z = e^{2\pi it}$ then we obtain a homeomorphism f on the unit circle X. For every $z \in X$, we have $\omega f(z) = \{1\}$, the unique fixed point. However, $\{1\}$ is not an attractor for f. In fact, $\Omega f(1) = X$.

The class of attractors is closed under finite union and finite intersection. Using infinite intersections we can characterize $\mathcal{C}f$ invariance.

Theorem 3.2 *Let f be a homeomorphism on X and A be a closed f invariant subset of X. The following conditions are equivalent. When they hold we call A a* quasi-attractor *for f.*

(i) A is $\mathcal{C}f$ + invariant.

(ii) A is the intersection of a (possibly infinite) set of attractors.

(iii) The set of inward neighborhoods of A form a basis for the neighborhood system of A.

From Theorem 2.3 again it follows that a quasi-attractor is stable.

If A is a closed invariant set for a homeomorphism f then A is called an *isolated invariant set* if it is the maximum invariant subset of some neighborhood U of A. That is,

$$A \quad = \quad \bigcap_{n=-\infty}^{+\infty} f^n(U). \tag{3.2}$$

In that case, U is called an *isolating neighborhood* for A. Notice that if U is a closed isolating neighborhood for A and the positive orbit of x remains in U, i. e. $f^n(x) \in U$ for $n = 0, 1, \ldots$ then $\omega f(x) \subseteq A$ because $\omega f(x)$ is an invariant subset of U.

Since an attractor is the maximum invariant subset of some inward set, it follows that an attractor is isolated. Conversely, by condition (iii) of Theorem 3.1 an invariant set is an attractor precisely when it is isolated and stable. In particular, a quasi-attractor is an attractor exactly when it is an isolated invariant set.

An attractor for f^{-1} is called a *repellor* for f. If U is an inward set for f then $\overline{X \setminus U} = X \setminus (U^\circ)$ is an inward set for f^{-1} and $B = \bigcap_{n=0}^{\infty} f^{-n}(\overline{X \setminus U})$ is the associated repellor for f. We call B the repellor *dual* to the attractor $A = \bigcap_{n=0}^{\infty} f^n(U)$. Recall that an f invariant set is f^{-1} invariant. In particular, attractors and repellors are both f and f^{-1} invariant. The open set $\bigcup_{n=0}^{\infty} f^{-n}(U) = X \setminus B$ is called the *domain of attraction* for A. The name comes from the implication:

$$\begin{aligned} x \in X \setminus B \quad &\Longrightarrow \quad \omega(x) \subseteq \Omega\mathcal{C}f(x) \subseteq A, \\ x \in X \setminus A \quad &\Longrightarrow \quad \alpha(x) \subseteq \Omega\mathcal{C}f^{-1}(x) \subseteq B. \end{aligned} \tag{3.3}$$

For example, the entire space X is both an attractor and a repellor with dual $\emptyset$.

If A is an attractor then we call the set of chain recurrent points in A, i. e. $A \cap |\mathcal{C}f|$, the *trace* of the attractor A. An attractor is determined by its trace via the equation

$$\mathcal{C}f(A \cap |\mathcal{C}f|) \quad = \quad A. \tag{3.4}$$

The trace of an attractor is a clopen subset of $|\mathcal{C}f|$ by part (vii) of Theorem 3.1. Conversely, suppose A_0 is a subset of $|\mathcal{C}f|$ which is + invariant for the restriction of $\mathcal{C}f$ to $|\mathcal{C}f|$. That is, $x \in A_0$ and $y \in \mathcal{C}f(x) \cap |\mathcal{C}f|$ implies $y \in A_0$. If A_0 is clopen in $|\mathcal{C}f|$, then $\mathcal{C}f(A_0)$ is the attractor with trace A_0 and $\mathcal{C}f^{-1}(|\mathcal{C}f| \setminus A_0)$ is the dual repellor. By Theorem 3.2, if A_0 is merely closed then $\mathcal{C}f(A_0)$ is a quasi-attractor.

With the relative topology the set $|\mathcal{C}f| = |\mathcal{C}f^{-1}|$ of chain recurrent points is a compact metric space and so has only countably many clopen subsets (Every clopen subset is a finite union of members of a countable basis for the topology). It follows that, while there are often uncountably many inward sets, there are only countably many attractors.

When restricted to $|\mathcal{C}f| = |\mathcal{C}f^{-1}|$ the closed relation $\mathcal{C}f \cap \mathcal{C}f^{-1}$ is reflexive as well as symmetric and transitive. The individual equivalence classes are closed f invariant subsets of X called the *chain components* of f, or the *basic sets* of f. These chain components are the analogues of the individual critical points in gradient case described in Section 1.

Any two points of a chain component are related by $\mathcal{C}f$. This is a type of transitivity condition. As with recurrence, there are several -increasingly broad- notions of dynamic transitivity and these can be associated with the relations of (2.18). First, we consider when the entire system f on X is transitive in the some way. Then we say that a closed f invariant subset A is a transitive subset in this way, when the subsystem $f|A$ on A is transitive in the appropriate way.

A group action on a set is called transitive when one can move from any element of the set to any other by some element of the group. That is, the entire set is a single orbit of the group action. Notice that this use of the word is unrelated to transitivity of a relation. Recall that a relation F on a set X is transitive when $F \circ F \subseteq F$. We are now considering when it happens that any two points of X are related by F. This just says that $F = X \times X$ which we will call the *total relation* on X.

First, what does it mean to say that $\mathcal{O}f$ is total, that is to say all of X lies in a single orbit of f? First, compactness implies that X is then finite and so the homeomorphism f is a permutation of the finite set X. Such permutation is a product of disjoint cycles and $\mathcal{O}f = X \times X$ exactly when all of X consists of a single cycle. The associated invariant subsets are the periodic orbits of f, including the fixed points.

Next, we consider when the orbit closure relation $\mathcal{R}f$ is total, that is, when every point is in the orbit closure of every other.

Theorem 3.3 *Let f be a homeomorphism on X. The following conditions are equivalent and when they hold we call f* minimal.

(i) For all $x \in X$, $\mathcal{O}f(x)$ is dense in X.

(ii) For all $x \in X$, $\mathcal{R}f(x) = X$, i. e. $\mathcal{R}f = X \times X$.

(iii) For all $x \in X$, $\omega f(x) = X$, i. e. $\omega f = X \times X$.

(iv) X is the only nonempty, closed $f+$ invariant subset of X.

(v) X is the only nonempty, closed f invariant subset of X.

Recall our convention that the state space of a dynamical system is nonempty, although we do allow the empty set as an invariant subset. For example, the empty set is the repellor/attractor dual to the attractor/repellor which is the entire space.

Thus, a closed f invariant subset A of X is *minimal* when it is nonempty but contains no nonempty, proper f invariant subset. From compactness it follows via the usual Zorn's Lemma argument that every nonempty, closed $f+$ invariant subset of X contains a minimal, nonempty, closed f invariant subset.

Theorem 3.4 *Let f be a homeomorphism on X. The following conditions are equivalent and when they hold we call f* topologically transitive.

(i) For some $x \in X$, $\mathcal{O}f(x)$ is dense in X, i. e. $\mathcal{R}f(x) = X$.

(ii) For all $x \in X$, $\mathcal{N}f(x) = X$, i. e. $\mathcal{N}f = X \times X$.

(iii) For all $x \in X$, $\Omega f(x) = X$, i. e. $\Omega f = X \times X$.

(iv) X is the only closed $f+$ invariant subset with a nonempty interior.

If f is topologically transitive then the set $\{x : \omega f(x) = X\}$ is residual, i. e. it is the countable intersection of dense, open subsets of X.

Every point in a minimal system has a dense orbit. If f is topologically transitive then the set of *transitive points* for f,

$$Trans_f \quad =_{def} \quad \{x : \omega f(x) = X\} \quad = \quad \{x : \mathcal{R}f(x) = X\}, \tag{3.5}$$

is residual and so is dense by the Baire Category Theorem. However, its complement is either empty, which is the minimal case, or else is dense as well. Also, there is a usually rich variety of invariant subsets. It may happen, for instance, that the set of periodic points is dense. Most well-studied examples of chaotic dynamical systems are non-minimal, topologically transitive systems. In fact, Devaney used the conjunction of topological transitivity and density of periodic points in an infinite system as a definition of chaos.

The broadest notion of transitivity is associated with the chain relation $\mathcal{C}f$.

Theorem 3.5 *Let f be a homeomorphism on X. The following conditions are equivalent and when they hold we call f* chain transitive.

(i) For all $x \in X$, $\mathcal{C}f(x) = X$, i. e. $\mathcal{C}f = X \times X$.

(ii) For all $x \in X$, $\Omega\mathcal{C}f(x) = X$, i. e. $\Omega\mathcal{C}f = X \times X$.

(iii) X is the only nonempty inward set.

(iv) X is the only nonempty attractor.

Before proceeding further, we pause to observe that each of these three concepts is the same for f and for its inverse. Because the f invariant subsets are the same as the f^{-1} invariant subsets, we see that f^{-1} is minimal when f is. Since $\mathcal{N}(f^{-1}) = (\mathcal{N}f)^{-1}$ and $\mathcal{C}(f^{-1}) = (\mathcal{C}f)^{-1}$, it follows as well that f^{-1} is topologically transitive or chain transitive when f satisfies the corresponding property.

There are many naturally occurring chain transitive subsets. For every $x \in X$ the limit sets $\alpha f(x)$ and $\omega f(x)$ are chain transitive subsets as are each of the chain components. Notice that if x and y are points of some chain component B then by definition of the equivalence relation $\mathcal{C}f \cap \mathcal{C}f^{-1}$, x and y can be connected by ϵ chains in X for any positive ϵ. To say that B is a chain transitive subset is to make the stronger statement that they can be connected via ϵ chains which remain in B.

Contrast this positive result with the trap set by implication (2.23) which suggests that $B = \omega f(x)$ is a topologically transitive subset. However, (2.23), which says $B \times B \subseteq \mathcal{N}f$, does not imply that the restriction $f|B$ is topologically transitive, i.e. $B \times B = \mathcal{N}(f|B)$. It may not be possible to get from near y to near z without a hop which takes you outside B.

In fact, if f is any chain transitive homeomorphism on a space X then it is possible to embed X as a closed subset of a space Y and extend f to a homeomorphism g on Y in such a way that $X = \omega g(y)$ for some point $y \in Y$.

Any chain transitive subset for f is contained in a unique chain component of $|\mathcal{C}f|$. In fact the chain components are precisely the maximal chain transitive subsets.

In particular, each subset $\alpha f(x)$ and $\omega f(x)$ is contained in some chain component of f.

By using (2.15) one can show that each connected component of the closed subset $|\mathcal{C}f|$ is contained in some chain component as well. It follows that the space of chain components, that is, the space of $\mathcal{C}f \cap \mathcal{C}f^{-1}$ equivalence classes with the quotient topology from $|\mathcal{C}f|$, is zero-dimensional. So this space is either countable or is the union of a Cantor set with a countable set.

An individual chain component B is called *isolated* when it is a clopen subset of the chain recurrent set $|\mathcal{C}f|$ or, equivalently, if it is an isolated point in the space of chain components. Thus, B is isolated when it has a neighborhood U in X such that $U \cap |\mathcal{C}f| = B$. It can be proved that B is an isolated chain component precisely when it is an isolated invariant set, i. e. it admits a neighborhood U satisfying (3.2).

The individual chain components generalize the role played by the critical points in the gradient case. They are the blobs we described at the end of Section 1. We complete the analogy by identifying which chain components are like the relative maxima and minima among the critical points.

Theorem 3.6 *Let B be a nonempty, closed, f invariant set for a homeomorphism f on X. The following conditions are equivalent and when they hold we call B a* terminal chain component.

(i) B is a chain transitive quasi-attractor.

(ii) *B is a $\mathcal{C}f$ + invariant, chain transitive subset of X.*

(iii) *B is a $\mathcal{C}f$ + invariant chain component.*

(iv) *In the set of nonempty, closed $\mathcal{C}f$ + invariant subsets of X B is an element which is minimal with respect to inclusion.*

In particular, B is a chain transitive attractor if and only if it is an isolated, terminal chain component.

From condition (iv) and Zorn's Lemma it follows that every nonempty, closed $\mathcal{C}f$ + invariant subset of X contains a terminal chain component. In particular, $\Omega\mathcal{C}f(x)$ contains a terminal chain component for every $x \in X$.

Clearly, if there are only finitely many chain components then each terminal chain component is isolated and so is an attractor.

Corollary 3.7 *Let f be a homeomorphism on X and let B be a nonempty, closed, $\mathcal{C}f$ + invariant subset of X. If for some $x \in X$, $B \subseteq \omega f(x)$ then B is a terminal chain component and $B = \omega f(x)$.*

Proof: B contains some terminal chain component B_1 and $\omega f(x)$ is contained in some chain component B_2. Thus, $B_1 \subseteq B \subseteq \omega f(x) \subseteq B_2$. Since distinct chain components are disjoint, $B_1 = B_2$.

□

Because $\mathcal{C}(f^{-1}) = (\mathcal{C}f)^{-1}$, the chain components for f and f^{-1} are the same. A chain component is called an *initial chain component* for f when it is a terminal chain component for f^{-1}.

Let L be a Lyapunov function for $\mathcal{C}f$. L is constant on each chain component and Theorem 2.5 says that L can be constructed to be strictly increasing on the orbit of every point of $X \setminus |\mathcal{C}f|$. That is, for such a complete Lyapunov function the set $|L|$ of critical points equals $|\mathcal{C}f|$. In that case, the local maxima occur at terminal chain components and, as a partial converse, an isolated terminal chain component is a local maximum for any complete $\mathcal{C}f$ Lyapunov function.

Theorem 3.8 *For a homeomorphism f on X let L be a $\mathcal{C}f$ Lyapunov function and let B be a chain component with r the constant value of L on B.*

(a) *If there exists a neighborhood U of B such that $L(x) < r$ for all $x \in U \setminus B$, then B is a terminal chain component.*

(b) *Assume L is a complete Lyapunov function. If B is an attractor then it is a terminal chain component and the domain of attraction U is an open set containing B such that $L(x) < r$ for all $x \in U \setminus B$.*

In particular, if there are only finitely many chain components for f then by using L, a complete $\mathcal{C}f$ Lyapunov function which distinguishes the chain components, we obtain the picture promised at the end of Section 1. The local maxima of L occur at the terminal chain components which are attractors. The local minima are at the initial chain components which are repellors. The remaining chain components play the role of saddles. As in the gradient case, it can happen that there is an open set of points x such that $\omega f(x)$ is contained in one of these saddle chain components. There is a natural topological condition which, when it holds, excludes this pathology.

Theorem 3.9 *For a homeomorphism f on X, assume that $\mathcal{N}f = \mathcal{C}f$ or, equivalently, that $\Omega f = \Omega\mathcal{C}f$.*

If A is a stable closed f invariant subset of X then A is a quasi-attractor.

For a residual set of points x in X, $\omega f(x)$ is a terminal chain component and $\alpha f(x)$ is an initial chain component.

Proof: The result for a stable set A follows from the characterizations in Theorems 2.3 and 3.2.

By Theorem 2.2, the set of x such that $\omega f(x) = \Omega f(x)$ is always residual. By assumption this agrees with the set of x such that $\omega f(x) = \Omega\mathcal{C}f(x)$. Corollary 3.7 implies that for such x the set $\omega f(x)$ is a terminal chain component. For $\alpha f(x)$ apply this result to f^{-1}. QED

If X is a compact manifold of dimension at least two, then the condition $\mathcal{N}f = \mathcal{C}f$ holds for a residual set in the Polish group of homeomorphisms on X with the uniform topology. If, in addition, f has only finitely many chain components (and this is not a residual condition on f) then it follows that the points which are in the domain of attraction of some terminal set and in the domain of repulsion of some initial set form a dense, open subset of X.

4 Chaos and Equicontinuity

Any chain transitive system can occur as a chain component, but in the most interesting cases the chain components are topologically transitive subsets. In this section we consider the antithetical phenomena of equicontinuity and chaos in topologically transitive systems.

If f is a homeomorphism on X and $x \in X$ then x is called a *transitive point* when its orbit $\mathcal{O}f(x)$ is dense in X. As in (3.4) we denote by $Trans_f$ the set of transitive points for f. The system is minimal exactly when every point is transitive, i. e. when $Trans_f = X$.

To study equicontinuity we introduce a new metric d_f defined using the original metric d on X.

$$d_f(x,y) \quad =_{def} \quad sup\ \{d(f^n(x), f^n(y)) : n = 0, 1, 2, ...\}. \tag{4.1}$$

It is easy to check that d_f is a metric, i. e. it satisfies the conditions of positivity and symmetry as well as the triangle inequality. However, it is usually not topologically equivalent to the original metric d, generating a topology which is usually strictly finer (= more open sets). We call a point $x \in X$ an *equicontinuity point* for f when for every $\epsilon > 0$ there exists a $\delta > 0$ so that for all $y \in X$

$$d(x,y) \ < \ \delta \qquad \Longrightarrow \qquad d_f(x,y) \ \leq \ \epsilon, \tag{4.2}$$

or, equivalently, if for every $\epsilon > 0$ there exists a neighborhood U of x with d_f diameter at most ϵ (the terms "neighborhood", "open set", etc. will always refer to the original topology given by d unless otherwise specified). Here the d_f diameter of a subset $A \subseteq X$ is

$$diam_f(A) \quad =_{def} \quad sup\ \{d_f(x,y) : x, y \in A\}. \tag{4.3}$$

We denote the, possibly empty, set of equicontinuity points for f by Eq_f.

When every point is an equicontinuity point, the system associated with f is called *equicontinuous*, or we just say that f is equicontinuous. This coincides with the concept of equicontinuity of the set of functions $\{f^n : n = 1, 2, ...\}$. Recall that any finite set of continuous functions is equicontinuous, but equicontinuity for an infinite set is often, as in this case, a strong condition. Equicontinuity of f says exactly that the metrics d and d_f are topologically equivalent. For compact spaces topologically equivalent metrics are uniformly equivalent. Hence, if f is equicontinuous, then for every $\epsilon > 0$ there exists a $\delta > 0$ such that for all $x, y \in X$ implication (4.2) holds.

When f is equicontinuous then we can replace d by d_f since in the equicontinuous case the latter is a metric giving the correct topology. The homeomorphism f is then an *isometry* of the metric, i. e. $d_f(x, y) = d_f(f(x), f(y))$ for all $x, y \in X$. This equality uses a famous little problem which keeps being rediscovered: If f is a surjective map on a compact metric space X with metric d such that $d(f(x), f(y)) \leq d(x, y)$ for every $x, y \in X$ then f is an isometry on X, see e. g. Alexopoulos [9] or Akin [3] Proposition 2.4(c).

Conversely, if f is an isometry of a metric d on X (with the correct topology) then f is clearly equicontinuous.

When X has a dense set of equicontinuity points we call the system *almost equicontinuous*. If f is almost equicontinuous but not equicontinuous then the δ in (4.2) will depend upon $x \in Eq_f$ as well as upon ϵ.

On the other hand, we say that the system *has sensitive dependence upon initial conditions*, or simply that f is *sensitive* when there exists $\epsilon > 0$ such that $diam_f(U) > \epsilon$ for every nonempty open subset U of X, or, equivalently, if there exists $\epsilon > 0$ such that for any $x \in X$ and $\delta > 0$, there exists $y \in X$ such that $d(x, y) < \delta$ but for some $n > 0$ $d(f^n(x), f^n(y)) > \epsilon$. Here the important issue is that ϵ is independent of the choice of x and δ.

Sensitivity is a popular candidate for a definition of chaos in the topological context. Suppose for a sensitive homeomorphism f you are attempting to estimate analyze the orbit of x, but may make an -arbitrarily small- positive error for initial point input. Even if you are able to compute the iterates exactly, then you cannot be certain that you always remain ϵ close to the orbit you want. If you have chosen a bad point y as input then at some time n you will be at $f^n(y)$ more than distance ϵ away from the point $f^n(x)$ that you want.

For transitive systems, the *Auslander-Yorke Dichotomy Theorem* holds (see [21]):

Theorem 4.1 *Let f be a topologically transitive homeomorphism on a compact metric space X. Exactly one of the following alternatives is true.*

- (Sensitivity) *The homeomorphism f is sensitive and there are no equicontinuity points, i. e. $Eq_f = \emptyset$.*
- (Almost Equicontinuity) *The set of equicontinuity points coincides with the set of transitive points, i. e. $Eq_f = Trans_f$, and so the set of*

equicontinuity points is residual in X.

Proof: For $\epsilon > 0$ let $Eq_{f,\epsilon}$ be the union of all open sets with d_f diameter at most ϵ. So $x \in Eq_{f,\epsilon}$ if and only if it has a neighborhood with d_f diameter at most ϵ. It is then easy to check that $x \in Eq_{f,\epsilon}$ if and only if $f(x) \in Eq_{f,\epsilon}$. Thus, $Eq_{f,\epsilon}$ is an f invariant open set. If $Eq_{f,\epsilon}$ is nonempty and x is a transitive point then the orbit eventually enters $Eq_{f,\epsilon}$. Since $f^n(x) \in Eq_{f,\epsilon}$ for some positive n, it follows that $x \in Eq_{f,\epsilon}$ because the set is invariant. This shows that

$$Eq_{f,\epsilon} \neq \emptyset \qquad \Longrightarrow \qquad Trans_f \subseteq Eq_{f,\epsilon}. \tag{4.4}$$

If for every $\epsilon > 0$ we have $Eq_{f,\epsilon} \neq \emptyset$ then $Trans_f \subseteq \bigcap_{\epsilon>0} Eq_{f,\epsilon} = Eq_f$. We omit the proof that only the transitive points can be equicontinuous in a topologically transitive system.

If, instead, for some $\epsilon > 0$ the set $Eq_{f,\epsilon}$ is empty then by definition f is sensitive. QED

The system is minimal if and only if $Trans_f = X$ and so a topologically transitive system is equicontinuous if and only if it is both almost equicontinuous and minimal. In particular, we have:

Corollary 4.2 *Let f be a minimal homeomorphism on X. Either f is sensitive or f is equicontinuous.*

Banks et al [22] showed that a topologically transitive homeomorphism on an infinite space with dense periodic points is always sensitive. On the other hand, there do exist topologically transitive homeomorphisms which are almost equicontinuous but not minimal and hence not equicontinuous. If f is an almost equicontinuous, topologically transitive homeomorphism then there is a sequence of positive integers $n_k \to \infty$ such that the sequence $\{f^{n_k}\}$ converges uniformly to the identity 1_X, see Glasner and Weiss [37] and Akin, Auslander and Berg [6]. In general, when such a convergent sequence of iterates exists the homeomorphism f is called *uniformly rigid*.

Other notions of chaos are defined using ideas related to proximality. A pair $(x, y) \in X \times X$ is called *proximal* for f if

$$lim\ inf_{n\to\infty}\, d(f^n(x), f^n(y)) \quad = \quad 0, \tag{4.5}$$

or, equivalently, if $1_X \cap \omega(f \times f)(x, y) \neq \emptyset$ where $f \times f$ is the homeomorphism on $X \times X$ defined by $(f \times f)(x, y) = (f(x), f(y))$. If a pair is not proximal then it is called *distal.* The pair is called *asymptotic* if

$$lim_{n \to \infty} \; d(f^n(x), f^n(y)) \quad = \quad 0. \tag{4.6}$$

We will call (x, y) a *Li-Yorke pair* if it is proximal but not asymptotic. That is, (4.5) holds but

$$lim \; sup_{n \to \infty} \; d(f^n(x), f^n(y)) \quad > \quad 0. \tag{4.7}$$

Li and Yorke [48] called a subset A of X a *scrambled set* if every non-diagonal pair in $A \times A$ is a Li-Yorke pair. Following their definition f is called *Li-Yorke chaotic* when there is an uncountable scrambled subset.

The homeomorphism f is called *distal* when every nondiagonal pair in $X \times X$ is a distal pair. If f is an isometry then clearly f is distal and so every equicontinuous homeomorphism is distal. There exist homeomorphisms f which are minimal and distal but not equicontinuous, as described in e. g. Auslander [17]. Since such an f is minimal but not equicontinuous, it is sensitive. On the other hand, a distal homeomorphism has no proximal pairs and so is certainly not Li-Yorke chaotic.

There exists an almost equicontinuous, topologically transitive, non-minimal homeomorphism f such that a fixed point $e \in X$ is the unique minimal subset of X and so the pair (e, e) is the unique subset in $X \times X$ which is minimal for $f \times f$. It follows that $(e, e) \in \omega(f \times f)(x, y)$ for every pair $(x, y) \in X \times X$ and so every pair is proximal. On the other hand, because f is uniformly rigid, every pair (x, y) is recurrent for $f \times f$ and it follows that no non-diagonal pair is asymptotic. Thus, the entire space X is scrambled and f is Li-Yorke chaotic but not sensitive.

There is a sharpening of topological transitivity which always implies sensitivity as well as most other topological conditions associated with chaos. A homeomorphism f on X is called *weak mixing* if the homeomorphism $f \times f$ on $X \times X$ is topologically transitive. If (x, y) is a transitive point for $f \times f$ then since $\omega(f \times f)(x, y) = X \times X$ it follows that $d_f(x, y) = M$, where M is the diameter of the entire space X. Suppose f is weak mixing and U is any nonempty open subset of X. Since $Trans_{f \times f}$ is dense in $X \times X$, there is a transitive pair (x, y) in $U \times U$. Hence, $diam_f(U) = M$. When f is weak mixing, there exists an uncountable $A \subseteq X$ such that every nondiagonal pair in $A \times A$ is a transitive point for $f \times f$, Iwanik [44] and Huang and Ye [42], see also Akin [5]. Since such a set A is clearly scrambled, it follows that f is Li-Yorke chaotic.

For any two subsets $U, V \subseteq X$ we define the *hitting time set* to be the set integers given by:

$$N(U, V) \quad =_{def} \quad \{n \geq 0 : f^n(U) \cap V \; \neq \; \emptyset\} \tag{4.8}$$

The topological transitivity condition $\Omega f = X \times X$ says that whenever U and V are nonempty open subsets the hitting time set $N(U, V)$ is infinite. We call f *mixing* if for every pair of nonempty open subsets U, V there exists k such that $n \in N(U, V)$ for all $n \geq k$. As the names suggest, mixing implies weak mixing.

Example The most important example of a chaotic homeomorphism is the ubiquitous *shift homeomorphism*. We think a of fixed finite set A as an *alphabet* and for any positive integer k the sequences in A of length k, i. e. the elements of A^k, are called *words of length* k. When A is equipped with the discrete topology it is compact and so by the Tychonoff Product Theorem any product of infinitely many copies of A is compact when given the product topology. Using the group of integers as our index set, we let $X = A^{\mathbb{Z}}$. There is a metric compatible with this topology. For $x, y \in X$ let

$$d(x, y) \quad =_{def} \quad infimum\ \{2^{-n} : x_i = y_i \text{ for all } i \text{ with } |i| < n\}. \tag{4.9}$$

The metric d is an *ultrametric*. That is, it satisfies a strengthening of the triangle inequality. For all $x, y, z \in X$:

$$d(x, z) \quad \leq \quad max(d(x, y), d(y, z)). \tag{4.10}$$

The ultrametric inequality is equivalent to the condition that for every positive ϵ the open relation V_ϵ is an equivalence relation.

For any word $a \in A^k$ and any integer j, the set $\{x \in X : x_{i+j} = a_i : i = 1, ..., k\}$ is a clopen subset of X called a *cylinder set*, which we will denote $U_{a,j}$. For example, with $k = 2n-1$ and $j = -n$ the cylinder sets are precisely the open balls of radius ϵ when $2^{-n} < \epsilon \leq 2^{-n+1}$. From this we see that X is a Cantor set with cylinder sets as a countable basis of clopen sets.

On X the shift homeomorphism s is defined by the equation

$$s(x)_i \quad = \quad x_{i+1} \qquad \text{for all } i \in \mathbb{Z}. \tag{4.11}$$

The fixed points $|s|$ are exactly the constant sequences in X and the periodic points $|\mathcal{O}s|$ are the periodic sequences in X. That is, $s^t(x) = x$ for some positive integer t exactly when $x_{i+t} = x_i$ for all $i \in \mathbb{Z}$.

It is easy to see that s is mixing. For example, if $a, b \in A^{2n-1}$ and $j = -n$, then the hitting time set $N(U_{a,j}, U_{b,j})$ contains every integer t larger than $2n$. For if $x_{i-n} = a_i$ and $x_{i-n+t} = b_i$ for $i = 1, ..., 2n-1$ then $x \in U_{a,j}$ and $s^t(x) \in U_{b,j}$.

This illustrates why s is chaotic in the sense of unpredictable. Given $x \in X$, if $y \in X$ satisfies $d(x, y) = 2^{-n}$, then moving right from position 0 the first n entries of y are known exactly. They agree with the entries of x. But after position n, x provides no information about y. The remaining entries on the right can be chosen arbitrarily.

Since s is mixing it is certainly topologically transitive, but it is useful to characterize the transitive points of s. For any point $x \in X$ and positive integer n we can scan from the central position x_0, left and right $n-1$ steps and observe a word of length $2n-1$. As we apply s the central position shifts right. When we have applied s^t it has shifted t steps. Scanning left and right we observe a new word. A point is a transitive point when for every n we can observe every word in A^{2n-1} in this way by varying t. To construct a transitive point x we need only list the -countably many- finite words of every length and lay them out end to end to get the right side of x. The left side of x can be arbitrary.

The shift homeomorphism is also expansive. In general, a homeomorphism f on a compact metric space X is called *expansive* when the diagonal 1_X is an isolated invariant set for the homeomorphism $f \times f$ on $X \times X$ and so there exists $\epsilon > 0$ such that $\bar{V}_\epsilon$ is an isolating neighborhood in the sense of (3.2). That is, if $x, y \in X$ and $d(f^n(x), f^n(y)) \leq \epsilon$ for all $n \in \mathbb{Z}$ then $x = y$. Such a number $\epsilon > 0$ is called an *expansivity constant*. With respect to the metric given by (4.9) it is easy to check that $\frac{1}{2}$ is an expansivity constant for the shift homeomorphism s.

The shift is interesting in its own right. In addition, the chaotic behavior of other important examples, especially expansive homeomorphisms, are studied by comparing them with the shift.

For a compact metric space X let $H(X)$ denote the automorphism group of X, i. e. the group of homeomorphisms on X, with the topology of uniform convergence. The metric on $H(X)$ is defined by the equation, for $f, g \in$

$H(X)$:

$$d(f,g) \quad =_{def} \quad supremum\{d(f(x),g(x)) : x \in X\}. \tag{4.12}$$

We obtain a topologically equivalent, but complete, metric by using $max(d(f,g), d(f^{-1}.g^{-1}))$. The automorphism group is a Polish (= admits a complete, separable metric) topological group. For $f \in H(X)$ we define the *translation* homeomorphisms ℓ_f and r_f on $H(X)$ by

$$\ell_f(g) \quad =_{def} \quad f \circ g \qquad \text{and} \qquad r_f(g) \quad =_{def} \quad g \circ f. \tag{4.13}$$

For any $x \in X$, the *evaluation map* $ev_x : H(X) \to X$ is the continuous map defined by $ev_x(f) = f(x)$.

For any $f \in H(X)$ let G_f denote the closure in $H(X)$ of the cyclic subgroup generated by f. That is,

$$G_f \quad = \quad \overline{\{f^n : n \in \mathbb{Z}\}} \quad \subseteq \quad H(X). \tag{4.14}$$

Thus, G_f is a closed, abelian subgroup of $H(X)$.

Let $Iso(X)$ denote the closed subgroup of isometries in $H(X)$ (In contrast with $H(X)$ this varies with the choice of metric). It follows from the Arzela-Ascoli Theorem that $Iso(X)$ is compact in $H(X)$. If f is an isometry on X then G_f is a compact, abelian subgroup of $Iso(X)$. Furthermore,the translation homeomorphisms ℓ_f and r_f restrict to isometries on the compact space $Iso(X)$ and G_f is a closed invariant subset. In fact, under either ℓ_f or r_f, G_f is just the orbit closure of f regarded as a point of $Iso(X)$.

Recall that if f is an equicontinuous homeomorphism, then we can replace the original metric by a topologically equivalent one, e. g. replace d by d_f, to get one for which f is an isometry.

If f is a homeomorphism on X and g is a homeomorphism on Y then we say that a continuous function $\pi : Y \to X$ *maps* g *to* f when $\pi \circ g = f \circ \pi : Y \to X$. If, in addition, π is a homeomorphism then we call π an *isomorphism* from g to f.

Theorem 4.3 *Let f be a homeomorphism on X. Fix $x \in X$. The evaluation map $ev_x : H(X) \to X$ maps ℓ_f on $H(X)$ to f on X.*

If f is an almost equicontinuous, topologically transitive homeomorphism then ev_x restricts to a homeomorphism from the closed subgroup G_f of $H(X)$ onto $Trans_f$ the residual subset of X consisting of the transitive points for f.

If f is a minimal isometry then ev_x restricts to a homeomorphism from the compact subgroup G_f of $H(X)$ onto X and so is an isomorphism from ℓ_f on G_f to f on X.

The isometry result is classical, see e. g. Gottschalk and Hedlund [38]. It shows that minimal, equicontinuous homeomorphisms are just translations on compact, monothetic groups, where a topological group is *monothetic* when it has a dense cyclic subgroup. Similarly a topologically transitive, almost equicontinuous homeomorphism which is not minimal is obtained by "compactifying" a translation on a noncompact monothetic group, see Akin and Glasner [7]. In fact, the group cannot even be locally compact and so, for example, is not the discrete group of integers. While examples of such topologically transitive, almost equicontinuous but not equicontinuous systems are known, it is not known whether there are any finite dimensional examples. And it is known that they cannot occur on a zero-dimensional space, i. e. the Cantor set. If finite dimensional examples do not exist then every topologically transitive homeomorphism on a compact manifold is sensitive except for the equicontinuous ones which we now describe.

Examples We can identify the circle S with the quotient topological group $\mathbb{R}/\mathbb{Z}$. For $a \in \mathbb{R}$ the translation $L_a = R_a$ on $\mathbb{R}$ induces the rotation τ_a on the circle $\mathbb{R}/\mathbb{Z}$. If $a \in \mathbb{Z}$ then this is the identity map. If a is rational then τ_a is periodic. But if a is irrational then τ_a is a minimal isometry on S. More generally, if $\{1, a_1, .., a_n\}$ is linearly independent over the field of rationals $\mathbb{Q}$ then on the torus $X = S^n$ the product homeomorphism $\tau_{a_1} \times ... \times \tau_{a_n}$ is a minimal isometry. Such systems are sometimes called *quasi-periodic.*

Of course, the translation by 1 on the finite cyclic group $\mathbb{Z}/k\mathbb{Z}$ of integers modulo k is just a version of a single periodic orbit, the unique minimal map on a finite space of cardinality k.

Recall that if $\{X_1, X_2, ...\}$ is a sequence of topological spaces and $\{p_n : X_{n+1} \to X_n\}$ is a sequence of continuous maps, then the *inverse limit* is the closed subset of the product space $\Pi_{n=1}^{\infty} X_n$:

$$LIM \ \{X_n, p_n\} \quad =_{def} \quad \{x : p_n(x_{n+1}) = x_n \text{ for } \ n = 1, 2, ...\}. \tag{4.15}$$

If the spaces are compact and the maps are surjective then the n^{th} coordinate projection π_n maps the compact space LIM onto X_n for every n (Hint: use Proposition 2.1). Also, if the spaces are topological groups and the maps are homomorphisms, then LIM is a closed subgroup of the product topological group. Finally, if $\{f_n : X_n \to X_n\}$ is a sequence of homeomorphisms such that p_n maps f_{n+1} to f_n for every n then the product homeomorphism $\Pi_{n=1}^{\infty} f_n$ restricts to a homeomorphism f on LIM and π_n maps f to f_n for every n.

Now let $\{k_n\}$ be an increasing sequence of positive integers such that k_n divides k_{n+1} for every n. Let X_n denote the finite cyclic group $\mathbb{Z}/k_n\mathbb{Z}$ and let $p_n : X_{n+1} \to X_n$ be the quotient homomorphism induced by the inclusion map $k_{n+1}\mathbb{Z} \to k_n\mathbb{Z}$. Let f_n denote the translation by 1 on X_n. Define X to be the inverse limit of this system with f the restriction to X of the product homeomorphism. X is a topological group whose underlying space is zero-dimensional and perfect, i. e. the Cantor set. The product homeomorphism is an isometry when we use the metric analogous to the one defined by (4.9). Furthermore, the restriction f to X is a minimal isometry. These systems are usually called *adding machines* or *odometers*. It can be proved that every equicontinuous minimal homeomorphism on a Cantor space is isomorphic to one of these. In general, every equicontinuous minimal homeomorphism is isomorphic to (1) a periodic orbit, (2) an adding machine, (3) a quasi-periodic motion on a torus, or (4) a product with each factor either an irrational rotation on a circle, an adding machine or a periodic orbit. The informal expression *strange attractor* is perhaps best defined as a topologically transitive attractor which is not equicontinuous, i. e. which is not one of these.

The word "chaos" suggests instability and unpredictability. However, for many examples what is most apparent is stability. For the Henon attractor or the Lorenz attractor the word "the" is used because in simulations one begins with virtually any initial point, performs the iterations and, after discarding an initial segment, one observes a particular, repeatable picture. The set as a whole is a predictable feature, stable under perturbation of initial point, an varying continuously with the defining parameters.

However, once the orbit is close enough to the attractor, essentially moving within the attractor itself, the motion is unpredictable, sensitive to arbitrarily small perturbations. What remains is statistical prediction. We cannot exactly predict when the orbit will enter some small subset of the attractor, but we can describe approximately the amount of time it spends in the subset. Such analysis requires an invariant measure and this takes us to the boundary between topological and measurable dynamics.

By a measure μ on a compact metric space X we will mean a Borel probability measure on the space. Such a measure acts, via integration, on the Banach algebra $\mathcal{C}(X)$ of continuous real-valued functions on X. The set $\mathcal{P}(X)$ of such measures can thus be regarded as a convex subset of the dual space of $\mathcal{C}(X)$. Inheriting the weak* topology on the dual space, $\mathcal{P}(X)$ becomes a compact, metrizable space. There is a natural inclusion map $\delta : X \to \mathcal{P}(X)$ which associates to $x \in X$ the point mass at x, denoted δ_x.

The *support* of a measure μ on X is the smallest closed set with measure 1. Denoted $|\mu|$ its complement can be obtained by taking the union out of a countable base for the topology of those members with measure 0. The measure has *full support* (or simply μ is *full*) when $|\mu| = X$. A measure is full when every nonempty open set has positive measure.

A continuous map $h : X \to Y$ induces a map $h_* : \mathcal{P}(X) \to \mathcal{P}(Y)$ which associates to μ the measure $h_*\mu$ defined by $h_*\mu(A) = \mu(h^{-1}(A))$ for every Borel subset A of Y. The continuous linear operator h_* is the dual of $h^* : \mathcal{C}(Y) \to \mathcal{C}(X)$ given by $u \mapsto u \circ h$. Furthermore, h_* is an extension of h. That is, $h_*(\delta_x) = \delta_{h(x)}$. The supports are related by

$$h(|\mu|) \quad = \quad |h_*\mu|. \tag{4.16}$$

To prove this, observe that for an open set $U \subset Y$

$$U \cap h(|\mu|) \ = \ \emptyset \qquad \Longleftrightarrow \qquad \mu(h^{-1}(U)) \ = \ 0 \qquad \Longleftrightarrow \qquad h_*\mu(U) \ = \ 0. \tag{4.17}$$

In particular, if f is a homeomorphism on X then f_* is a linear homeomorphism on $\mathcal{P}(X)$. extending f on X. A measure μ is called an *invariant measure* for f. Thus, $|f_*|$ the set of fixed points for f_* is the set of invariant measures for f. Clearly, $|f_*|$ is compact, convex subset of $\mathcal{P}(X)$. The classical theorem of Krylov and Bogolubov says that this set is nonempty. To prove it, one begins with an arbitrary point $x \in X$ and considers the sequence of Cesaro averages

$$\sigma_n(f, x) \quad =_{def} \quad \frac{1}{n+1}\Sigma_{i=0}^{n}\ \delta_{f^n(x)}. \tag{4.18}$$

It can be shown that every measure in the nonempty set of limit points of this sequence lies in $|f_*|$. If the set of limit points consists of a single measure μ, i. e. the sequence converges to μ, then x is called a *convergence point* for the invariant measure μ. For $\mu \in |f_*|$ we denote by $Con(\mu)$ the -possibly empty- set of convergence points for μ.

The extreme points of the compact convex set $|f_*|$ are called the *ergodic* measures for f. That is, μ is ergodic if it is not in the interior of some line segment connecting a pair of distinct invariant measures. Equivalently, μ is ergodic if a Borel subset A of X is invariant, i. e. $f^{-1}(A) = A$, only when the measure $\mu(A)$ is either 0 or 1. It follows that if $u : X \to \mathbb{R}$ is a Borel measurable, invariant function, then for any ergodic measure μ there is a set of measure 1 on which u is constant. Or more simply, $u \circ f = u$ implies u is constant almost everywhere with respect to μ.

The central result in measurable dynamics is the Birkhoff Pointwise Ergodic Theorem which says, in this context:

Theorem 4.4 *Given a homeomorphism f on a compact metric space X, let $u : X \to \mathbb{R}$ be a bounded, Borel measurable function. There exists a bounded, invariant, Borel measurable function. $\hat{u} : X \to \mathbb{R}$ such that for every f invariant measure μ*

$$\int u \, d\mu \quad = \quad \int \hat{u} \, d\mu \qquad \textit{and} \qquad lim_{n \to \infty} \frac{1}{n+1} \Sigma_{i=0}^{n} \, u(f^n(x)) \; = \; \hat{u}(x) \tag{4.19}$$

almost everywhere with respect to μ.

In particular, if μ is ergodic then

$$\hat{u}(x) \quad = \quad \int u \, d\mu \tag{4.20}$$

almost everywhere with respect to μ.

It is a consequence of the ergodic theorem that

$$\mu \;\text{ ergodic} \qquad \Longrightarrow \qquad \mu(Con(\mu)) \; = \; 1. \tag{4.21}$$

Thus, with respect to an ergodic measure μ for almost every point x,

$$lim_{n \to \infty} \frac{1}{n+1} \Sigma_{i=0}^{n} \, u(f^n(x)) \quad = \quad \int u \, d\mu, \tag{4.22}$$

for every $u \in \mathcal{C}(X)$. The left hand side is the time-average of the function u along the orbit with initial point x and it equals the space-average on the right.

For any f invariant measure μ it follows from (4.16) that the support, $|\mu|$, is an f invariant subset of X. The Poincaré Recurrence Theorem says that if U is any open set with $U \cap |\mu| \neq \emptyset$ then U is non-wandering. That is, for some positive integer n, $U \cap f^{-n}(U) \neq \emptyset$. This is clear from the observation that the open sets $f^{-n}(U)$ all have the same, positive measure and so they cannot all be pairwise disjoint. Applying this to the subsystem obtained by restricting to $|\mu|$ it follows from Theorem 2.2 that the set of recurrent points in $|\mu|$ form a dense G_δ subset of $|\mu|$.

If $x \in Con(\mu)$ then the orbit of x is dense in μ and so the subsystem on $|\mu|$ is topologically transitive whenever $Con(\mu)$ is nonempty. By the Birkhoff Ergodic Theorem this applies whenever μ is ergodic.

Since the support is a nonempty, closed, invariant subspace it follows that if f is minimal then every invariant measure has full support. The homeomorphism f is called *strictly ergodic* when it is minimal and has a unique invariant measure μ, which is necessarily ergodic. In that case, it can be shown that every point is a convergence point for μ, that is, $Con(\mu) = X$.

We note that in the dynamical systems context the topological notion of residual (that is, a dense G_δ subset) is quite different from the measure theoretic idea (a set of full measure). For example,

$$Con \quad =_{def} \quad \bigcup_{\mu \in |f_*|} Con(\mu) \tag{4.23}$$

is the set of points whose associated Cesaro average sequence is a Cauchy sequence. It follows that Con is a Borel set. By (4.21) $\mu(Con) = 1$ for every ergodic measure μ. Since every invariant measure is a limit of convex combinations of ergodic measures it follows that Con has measure 1 for every invariant measure μ.

On the other hand, it can be shown that if f is the shift homeomorphism on $X = A^{\mathbb{Z}}$ then for x in the dense G_δ subset of X the set of limit points of the sequence $\{\sigma_n(f, x)\}$ is all of $|f_*|$, Denker et al [30] or Akin [1] Chapter 9. Since the shift has many different invariant measures it follows that Con is disjoint from this residual subset. In general, if a homeomorphism f has more than one full, ergodic measure then Con is of first category, Akin [1] Theorem 8.11. In particular, if f is minimal but not strictly ergodic then Con is of first category.

The plethora of invariant measures undercuts somewhat their utility for statistical analysis. Suppose that there are two different ergodic measures μ and ν with common support, some invariant subset A of X. By restricting to A, we can reduce to the case when $A = X$ and so μ and ν are distinct full, ergodic measures. The sets $Con(\mu)$ and $Con(\nu)$ are disjoint and of measure 1 with respect to μ and ν, respectively. For an open set U the average frequency with which an orbit of $x \in Con(\mu)$ lies is U is given by $\mu(U)$ and similarly for ν. These mutually singular measures lead to different statistics.

One solution to this difficulty is to select what seems to be the best invariant measure in some sense, e. g. the measure of maximum entropy (see the article by King) if it should happen to be unique. However, as our introductory discussion illustrates, this somewhat misses the point.

Return to the case of a chaotic attractor A, a closed invariant subset of X. It often happens that the state space X comes equipped with a natural measure λ or at least a Radon-Nikodym equivalence class of measures, all with the same sets of measure 0. For example, if X is a manifold then λ is locally Lebesgue measure. The measure λ is usually not f invariant and it often happens that the set A of interest has λ measure 0. What we want, an *appropriate* measure μ for this situation, would be an invariant measure with support A, i. e.

$$\mu \in |f_*| \qquad \text{and} \qquad |\mu| = A, \tag{4.24}$$

and an open set U containing A such that with respect to λ almost every point of U is a convergence point for μ. That is,

$$\lambda(U \setminus Con(\mu)) = 0. \tag{4.25}$$

Notice that for $x \in Con(\mu)$, $\omega f(x) = |\mu| = A$ and so by Corollary 3.7 A is a terminal chain component. That is, for such a measure to exist, the attractor A must be at least chain transitive. If, in addition, $Con(\mu) \cap A \neq \emptyset$ then A is topologically transitive.

At least when strong hyperbolicity conditions hold this program can be carried out with μ the Bowen measure for the invariant set, see Katok-Hasselblatt [45] Chapter 20.

5 Minimality and Multiple Recurrence

In this section we provide a sketch of some important topics which were neglected in the above exposition. We first consider the study of minimal systems.

In Theorem 3.3 we defined a homeomorphism f on a compact space X to be *minimal* when every point has a dense orbit. A subset A of X is a minimal subset when it is a nonempty, closed, invariant subset such that the restriction $f|A$ defines a minimal homeomorphism on A. The term "minimal" is used because $f|A$ is minimal precisely when A is minimal, with respect to inclusion, in the family of nonempty, closed, invariant subsets. By Zorn's lemma every such subset contains a minimal subset. Since every system contains minimal systems, the classification of minimal systems provides a foundation upon which to build an understanding of dynamical systems in general.

On the other hand, if you start with the space, homeomorphisms which are minimal -on the whole space- are rather hard to construct. Some spaces have the fixed-point property, i. e. every homeomorphism on the space has a fixed point, and so admit no minimal homeomorphisms. The tori, which are monothetic groups, admit the equicontinuous minimal homeomorphisms described in the previous section. Fathi and Herman [34] constructed a minimal homeomorphism on the 3-sphere. For most other connected, compact manifolds it is not known whether they admit minimal homeomorphisms or not. It is even difficult to construct topologically transitive homeomorphisms, but for these a beautiful -but nonconstructive- argument due to Oxtoby shows that every such manifold admits topologically transitive homeomorphisms. He uses that Baire Category Theorem to show that if the dimension is at least two then the topologically transitive homeomorphisms are residual in the class of volume preserving homeomorphisms, see [54] Chapter 18.

Since we understand the equicontinuous minimal systems, we begin by building upon them. This requires a change in out point of view. Up to now we have mostly considered the behavior of a single dynamical system (X, f) consisting of a homeomorphism f on a compact metric space X. Regarding these as our objects of study we turn now to the maps between such systems. A *homomorphism* of dynamical systems, also called an *action map*, $\pi : (X, f) \to (Y, g)$ is a continuous map $\pi : X \to Y$ such that $g \circ \pi = \pi \circ f$ and so, inductively, $g^n \circ \pi = \pi \circ f^n$ for all $n \in \mathbb{Z}$. Thus, π maps the orbit of a point $x \in X$ to the orbit of $\pi(x) \in Y$. In general, for the map $\pi \times \pi : X \times X \to Y \times Y$ we have that

$$\begin{aligned} \pi \times \pi(\mathcal{A}f) \quad &\subseteq \quad \mathcal{A}g \qquad \text{for } \ \mathcal{A} = \mathcal{O}, \mathcal{R}, \mathcal{N}, \mathcal{G}, \mathcal{C} \\ \pi \times \pi(\Theta f) \quad &\subseteq \quad \Theta g \qquad \text{for } \ \Theta = \omega, \alpha, \Omega, \Omega\mathcal{C}. \end{aligned} \tag{5.1}$$

That is, π maps the various dynamic relations associated with f to the corresponding relations for g.

If π is bijective then it is called an *isomorphism* between the two systems and the inverse map π^{-1} is an action map, continuous by compactness.

If X is a closed invariant subset of Y with $f = g|X$ then the inclusion map π is an action map and then (X, f) is called the *subsystem* of (Y, g) determined by the invariant set X.

On the other hand, if π is surjective then (Y, g) is called a *factor*, or *quotient system*, of (X, f) and (X, f) is called a *lift* of (Y, g). A surjective action map is called a *factor map*.

For a factor map $\pi : (X, f) \to (Y, g)$ we define an important subset $R(\pi)$ of $X \times X$:

$$R(\pi) \quad =_{def} \quad \{(x_1, x_2) \in X \times X : \pi(x_1) = \pi(x_2)\} \quad = \quad (\pi \times \pi)^{-1}(1_Y). \tag{5.2}$$

The subset $R(\pi)$ is an ICER on X. That is, it is an invariant, closed equivalence relation. In general, if R is any ICER on X then on the space of equivalence classes the homeomorphism f induces a homeomorphism and the natural quotient map is a factor map of dynamical systems. The original factor system (Y, g) is isomorphic to the quotient system obtained from the ICER $R(\pi)$.

Notice that if (Y, g) is the *trivial system*, meaning that Y consisting of a single point, then $R(\pi)$ is the total relation $X \times X$ on X.

We use the ICER $R(\pi)$ to extend various definitions from dynamical systems to action maps between dynamical systems. For example, a factor map $\pi : (X, f) \to (Y, g)$ is called *equicontinuous* if for every $\epsilon > 0$ there exists $\delta > 0$ such that

$$(x_1, x_2) \in R(\pi) \ \text{ and } \ d(x_1, x_2) < \delta \qquad \Longrightarrow \qquad d_f(x_1, x_2) < \epsilon. \tag{5.3}$$

Comparing this with (4.2) we see that (X, f) is equicontinuous if and only if the factor map to the trivial system is equicontinuous.

Similarly, recall that $(x_1, x_2) \in X$ is a *distal pair* for (X, f) if $\omega(f \times f)(x_1, x_2)$ is disjoint from the diagonal 1_X, and the system (X, f) is *distal* when every nondiagonal pair is distal. A factor map $\pi : (X, f) \to (Y, g)$ is distal when every nondiagonal pair in $R(\pi)$ is distal. Again (X, f) is a distal system if and only if the factor map to the trivial system is distal.

It is easy to check that a distal lift of a distal system is distal. Since an equicontinuous factor map is distal, it follows that an equicontinuous lift of an equicontinuous system is distal. However, it need not be equicontinuous.

Example: If a is irrational then the rotation τ_a on the circle $Y = \mathbb{R}/\mathbb{Z}$, defined by $x \mapsto a + x$, is an equicontinuous minimal map. On the torus $X = \mathbb{R}/\mathbb{Z} \times \mathbb{R}/\mathbb{Z}$ we define f by

$$f(x, y) \quad =_{def} \quad (a + x, x + y). \tag{5.4}$$

It can be shown that (X, f) is minimal. The projection π to the first coordinate defines an equicontinuous factor map to $Y, \tau_a)$ and so (X, f) is distal. It is not, however, equicontinuous.

The above projection map is an example of a *group extension.* Let G be a compact topological group, like $\mathbb{R}/\mathbb{Z}$. Given any dynamical system (Y, g) and any continuous map $q : Y \to G$ we let $X = Y \times G$ and define f on SX by:

$$f(x, y) \quad =_{def} \quad (g(x), L_{q(x)}(y)). \tag{5.5}$$

The homeomorphism commutes with $1_Y \times R_z$ for any group element z and from this it easily follows that the projection π to the first coordinate is an equicontinuous factor map. If (Y, g) is minimal then the restriction of π to any minimal subset of X defines an equicontinuous factor map from the associated minimal subsystem. It can be shown that any equicontinuous factor map between minimal systems can be obtained via a factor from such a group extension.

The Furstenberg Structure Theorem says that any distal minimal system can be obtained by a -possibly transfinite- inverse limit construction, beginning with an equicontinuous system and such that each lift is an equicontinuous factor map. For the details of this and the structure theorem due to Veech for more general minimal systems we refer the reader to Auslander [17] Chapters 7 and 14, respectively.

The factor maps described above are projections from products. There exist examples which are not product projections even locally. For example, in Auslander [17] Chapter 1 the author uses a construction due to Floyd to build an action map π between minimal systems which is not an isomorphism but which is *almost one-to-one*. That is, for a residual set of points x in the domain the set of preimages $\pi^{-1}(\pi(x))$ is a singleton. Such a factor map is the opposite of distal. It is a *proximal* map, meaning that every pair $(x_1, x_2) \in R(\pi)$ is a proximal pair.

Also, one cannot base all one's constructions upon equicontinuous minimal systems. A dynamical system (X, f) is called *weak mixing* when the product $(X \times X, f \times f)$ is topologically transitive. Inclusion (5.1) implies that if a dynamical system is minimal, topologically transitive or chain transitive then any factor satisfies the corresponding property. It follows that any factor of a weak mixing system is weak mixing. It is clear that only the trivial system is both weak mixing and equicontinuous. Hence for a weak mixing system the trivial system is the only equicontinuous factor. For a minimal system the converse is true: if the trivial system is the only equicontinuous factor then the system is weak mixing. Furthermore, nontrivial weak mixing, minimal systems do exist.

Gottschalk and Hedlund in [38] introduced the idea using various special families of subsets of $\mathbb{N}$, the set of nonnegative integers, in order to distinguish different sorts of recurrence. A *family* $\mathcal{F}$ is a collection of subsets of $\mathbb{N}$ which is hereditary upwards. That is, if $A \subseteq B$ and $A \in \mathcal{F}$ then $B \in \mathcal{F}$. The family is called *proper* when it is a proper subset of the entire power set of $\mathbb{N}$, i. e. it is neither empty nor the entire power set. The heredity property implies that a family $\mathcal{F}$ is proper iff $\mathbb{N} \in \mathcal{F}$ and $\emptyset \not\in \mathcal{F}$. For a family $\mathcal{F}$ the dual family, denoted $\mathcal{F}^*$ (or sometimes $k\mathcal{F}$) is defined by:

$$\mathcal{F}^* \quad =_{def} \quad \{B \subseteq \mathbb{N} : B \cap A \neq \emptyset \text{ for all } A \in \mathcal{F}\} \quad = \quad \{B \subseteq \mathbb{N} : \mathbb{N} \setminus B \not\in \mathcal{F}\}. \tag{5.6}$$

It is easy to check that $\mathcal{F}^{**} = \mathcal{F}$ and that $\mathcal{F}^*$ is proper if and only if $\mathcal{F}$ is.

A *filter* is a proper family which is closed under pairwise intersection. A family $\mathcal{F}$ is the dual of a filter when it satisfies what Furstenberg calls the *Ramsey Property*:

$$A \cup B \in \mathcal{F} \qquad \Longrightarrow \qquad A \in \mathcal{F} \quad \text{or} \quad B \in \mathcal{F}. \tag{5.7}$$

For example, a set is in the dual of the family of infinite sets if and only if it is cofinite, i. e. its complement is finite. The family of cofinite sets is a filter.

A subset $A \subseteq \mathbb{N}$ is called *thick* when it contains arbitrarily long runs. That is, for every positive integer L there exists i such that $i, i+1, ..., i+L \in A$. Dual to the thick sets are the syndetic sets. A subset B is called *syndetic* or *relatively dense* if there exists a positive integer L such that every run of length L meets B.

In (4.8) we defined the hitting time set $N(U, V)$ for (X, f) when $U, V \subseteq X$. When U is a singleton $\{x\}$ we omit the braces and so have

$$N(x, V) \quad = \quad \{n \geq 0 : f^n(x) \in V\}. \tag{5.8}$$

It is clear that (X, f) is topologically transitive when $N(U, V)$ is nonempty for every pair of nonempty open sets U, V. Furstenberg showed that the stronger property that (X, f) be weak mixing is characterized by the condition that each such $N(U, V)$ is thick.

Recall that a point $x \in X$ is recurrent when $x \in \omega f(x)$. The point is called a *minimal point* when it is an element of a minimal subset of X in which case this minimal subset is $\omega f(x)$. Clearly, point $x \in X$ is recurrent when $N(x, U)$ is nonempty for every neighborhood U of x. Gottschalk and Hedlund proved that x is a minimal point if and only if every such $N(x, U)$ is relatively dense. For this reason, minimal points are also called *almost periodic* points.

Furstenberg reversed this procedure by using dynamical systems arguments to derive properties about families of sets and more generally to derive results in combinatorial number theory.

If $f_1, ..., f_k$ are homeomorphisms on a space X then $x \in X$ is a *multiple recurrent point* for $f_1, ..., f_k$ if there exists a sequence of positive integers $n_i \to \infty$ such that the k sequences $\{f_1^{n_i}(x)\}, ..., \{f_k^{n_i}\}$ all have limit x. That is, the point $(x, ..., x)$ is a recurrent point for the homeomorphism $f_1 \times ... \times f_k$ on X^k. The Furstenberg Multiple Recurrence Theorem says:

Theorem 5.1 *If $f_1, ..., f_k$ are commuting homeomorphisms on a compact metric space X, i. e. $f_i \circ f_j = f_j \circ f_i$ for $i, j = 1, ..., k$, then there exists a multiple recurrent point for $f_1, ..., f_k$.*

Corollary 5.2 *If f is a homeomorphism on a compact metric space X and $x \in X$ then for every positive integer k and every $\epsilon > 0$ there exist positive integers m, n such that with $y = f^m(x)$ the distance between any two of the points $y, f^n(y), f^{2n}(y), ..., f^{kn}(y)$ is less than ϵ.*

Proof: This follows easily by applying the Multiple Recurrence Theorem to the restrictions of the homeomorphisms $f, f^2, ..., f^k$ to the closed invariant subset $\omega f(x)$. Obtain a multiple recurrent point $y' \in \omega f(x)$ and then given $\epsilon > 0$ choose $y = f^m(x)$ to approximate y' sufficiently closely.
$\square$

These results are of great interest in themselves and they have been extended in various directions, see Bergelson-Leibman [24] for example. In addition, Furstenberg used the corollary to obtain a proof of Van der Waerden's Theorem:

Theorem 5.3 *If $B_1, ..., B_p$ is a partition of $\mathbb{N}$ then at least one of these sets contains arithmetic progressions of arbitrary length.*

Proof It suffices to show that for each $k = 1, 2, ..$ and some $a = 1, .., p$ the subset B_a contains an arithmetic progression of length $k + 1$. For this will then apply to some fixed B_a for infinitely many k. Let $A = \{1, ..., p\}$ and on $X = A^{\mathbb{Z}}$ define the shift homeomorphism defined by (4.11). Using the metric given by (4.9) and $\epsilon \leq \frac{1}{2}$ we observe that if $x, y \in X$ with $d(x, y) < \epsilon$ then $x_0 = y_0$.

Choose $x \in X$ such that $x_i = a$ if and only if $i \in B_a$ for $i \in \mathbb{N}$. Apply Corollary 5.2 to x and ϵ. Choosing positive integers m, n such that the points $f^m(x), f^{m+n}(x), f^{m+2n}(x), ..., f^{m+kn}(x)$ all lie within ϵ of each other we have that $m, m + n, ..., m + kn$ all lie in B_a where a is the common value of the 0 coordinate of these points.

□

One of the great triumphs of modern dynamical systems theory is Furstenberg's use of the ergodic theory version of these arguments to prove Szemerédi's Theorem, [35] Theorem 3.21.

Theorem 5.4 *Let B be a subset of $\mathbb{N}$ with positive upper Banach density, that is*

$$limsup_{|I| \to \infty} \ |B \cap I| \quad > \quad 0, \tag{5.9}$$

where I varies over bounded subintervals of $\mathbb{N}$ and $|A|$ denotes the cardinality of $A \subset \mathbb{N}$. The set B contains arithmetic progressions of arbitrary length.

For details and further results, we refer to Furstenberg's beautiful book [35].

We conclude by observing that our restriction throughout to the dynamics of homeomorphisms was a matter of expository convenience. Most of the definitions and results extend to the case where $f : X \to X$ is a surjective continuous map. One way of extending the theory is to use the *natural extension* of f. This is obtained by using a sequence $\{X_n\}$ of copies of X and

$p_n = f$ for all n. On the inverse limit space $\tilde{X}$ we define the homeomorphism $\tilde{f}$ by

$$\tilde{f}(x)_n \quad =_{def} \quad \begin{cases} f(x_0) & \text{for } n = 0, \\ x_{n-1} & \text{for } n > 0. \end{cases} \tag{5.10}$$

Many dynamic properties hold for the map f on X if and only if they hold for the homeomorphism $\tilde{f}$ on $\tilde{X}$.

6 Additional Reading

The approach to topological dynamics which was described in Sections 2 and 3 is presented in detail in Akin [1] and [2]. For the deeper, more specialized work in topological dynamics see Auslander [17], Ellis [33] and Akin [4]. A general survey of the field is given in de Vries [66]. Furstenberg [35] is a classic illustration of the applicability of topological dynamics to other fields.

Clark Robinson's text [56] is an excellent introduction to dynamical systems in general. More elementary introductions are Devaney [31] and Alligood et al [10]. Hirsch et al [40] provides a nice transition from differential equations to the general theory.

Much of modern differentiable dynamics grows out of the work of Smale and his students, see especially the classic Smale [61], included in the collection [62]. The seminal paper Shub and Smale [59] provides a bridge between this work and the purely topological aspects of attractor theory, see also Shub [58].

The fashionable topic of chaos has generated a large sample of writing whose quality exhibits extremely high variance. For expository surveys I recommend Lorenz [49] and Stewart [63]. An excellent collection of relatively readable, classic papers is Hunt et al [43].

For applications in biology see Hofbauer and Sigmund [41] and May [50]. Also, don't miss Sigmund's delightful book [60].

7 Bibliography

[1] Akin E (1993) The General Topology of Dynamical Systems, Amer. Math. Soc.,Providence, RI

[2] Akin E (1996) Dynamical systems: the topological foundations, in: Aulbach B and Colonius F, (eds) Six Lectures on Dynamical Systems, World Scientific, Singapore, pp. 1-43.

[3] Akin E (1996) On chain continuity, Discrete and Continuous Dynamical Systems 2:111-120

[4] Akin E (1997) Recurrence in Topological Dynamics: Furstenberg Families and Ellis Actions, Plenum Press, New York

[5] Akin E (2004) Lectures on Cantor and Mycielski sets for dynamical systems, in: Assani I, (ed) Chapel Hill Ergodic Theory Workshops, Amer. Math. Soc., Providence, pp. 21-80

[6] Akin E, Auslander J and Berg K (1996) When is a transitive map chaotic?, in: Bergelson V, March K and Rosenblatt J (eds) Conference in Ergodic Theory and Probability, Walter de Gruyter, Berlin, pp. 25-40

[7] Akin E and Glasner E (2001) Residual properties and almost equicontinuity, J d'Analyse Math 84:243-286

[8] Akin E, Hurley M and Kennedy, JA (2003) Dynamics of Topologically Generic Homeomorphisms, Memoir 783, Amer. Math. Soc.,Providence, RI

[9] Alexopoulos J (1991) Contraction mappings in a compact metric space, Solution No.6611, Math. Assoc. Amer. Monthly 98:450

[10] Alligood KT, Sauer TD and Yorke JA (1996) Chaos: An Introduction to Dynamical Systems, Spring Science- Business Media, New York

[11] Almgren FJ (1966) Plateau's Problem, W. A. Benjamin, Inc., New York

[12] Alpern SR and Prasad VS (2001) Typical Dynamics of Volume Preserving Homeomorphisms, Cambridge Univ. Press, Cambridge

[13] Arhangel'skii AV and Pontryagin LS (eds.)(1990) General topology I, Springer-Verlag, Berlin

[14] Arhangel'skii AV (ed.)(1995) General Topology III, Springer-Verlag, Berlin

[15] Arhangel'skii AV (ed.)(1996) General Topology II, Springer-Verlag, Berlin

[16] Auslander J (1964) Generalized recurrence in dynamical systems, in: Contributions to Differential Equations vol. 3, John Wiley, New York, pp. 55-74

[17] Auslander J (1988) Minimal Flows and Their Extensions, North-Holland, Amsterdam

[18] Auslander J, Bhatia N and Siebert P (1964) Attractors in dynamical systems, Bol. Soc. Mat. Mex. 9:55-66

[19] Auslander J and Siebert P (1963) Prolongations and generalized Liapunov functions in: International Symposium on Nonlinear Differential Equations and Nonlinear Mechanics, Academic Press, New York, pp. 454-462

[20] Auslander J and Siebert P (1964) Prolongations and stability in dynamical systems, Ann. Inst. Fourier, Grenoble 14, 2:237-268

[21] Auslander J and Yorke J (1980) Interval maps, factors of maps and chaos, Tohoku Math. J. 32:177-188

[22] Banks J, Brooks J, Cairns G, Davis G and Stacey P (1992) On Devaney's Definition of Chaos, Amer. Math. Monthly 99: 332-333

[23] Barrow-Green J (1997) Poincaré and the Three Body Problem, Amer. Math. Soc., Providence

[24] Bergelson V and Leibman A (1996) Polynomial extensions of Van der Waerden's and Szemeredi's theorems, J. Amer. Math. Soc. 9: 725-753

[25] Bing RH (1988) The Collected Papers of R. H. Bing, Amer. Math. Soc., Providence

[26] Brown M (ed) (1991) Continuum Theory and Dynamical Systems, Amer. Math. Soc., Providence

[27] Bhatia N and Szego G (1970) Stability Theory of Dynamical Systems, Springer-Verlag, Berlin

[28] Coddington E and Levinson N (1955) Theory of Ordinary Differential Equations, McGraw-Hill, New York

[29] Conley C (1978) Isolated Invariant Sets and the Morse Index, Amer. Math. Soc., Providence, RI

[30] Denker M, Grillenberger C and Sigmund K (1978) Ergodic Theory on Compact Spaces, Lect. Notes in Math 527, Springer-Verlag, Berlin

[31] Devaney RL (1989) An Introduction to Chaotic Dynamical Systems (2^{nd} ed), Addison-Wesley Publishing Company, Redwood City

[32] Douglas J (1939) Minimal surfaces of higher topological structure, Annals of Math.40:205-298

[33] Ellis R (1969) Lectures on Topological Dynamics, W A Benjamin Inc, New York

[34] Fathi A and Herman (1977) Existence de difféomorphismes minimaux, Soc. Math. de France, Astérique 49:37-59

[35] Furstenberg H (1981) Recurrence in Ergodic Theory and Combinatorial Number Theory, Princeton Univ. Press, Princeton

[36] Glasner E (2003) Ergodic Theory Via Joinings, Amer. Math. Soc., Providence, RI

[37] Glasner E and Weiss B (1993) Sensitive dependence on initial conditions, Nonlinearity 6:1067-1075

[38] Gottschalk W and Hedlund G (1955) Topological Dynamics, Amer. Math. Soc., Providence, RI

[39] Hartman (1964) Ordinary Differential Equations, John Wiley, New York

[40] Hirsch MW, Smale S and Devaney RL (2004) Differential Equations, Dynamical Systems and an Introduction to Chaos (2^{nd} ed) Elsevier Academic Press, Amsterdam

[41] Hofbauer J and Sigmund K (1988) The Theory of Evolution and Dynamical Systems, Cambridge Univ. Press, Cambridge

[42] Huang W and Ye X (2002) Devaney's chaos or 2-scattering implies Li-Yorke's chaos, Topology Appl. 117:259-272

[43] Hunt B, Kennedy JA, Li T-Y and Nusse H (eds.) (2004) The Theory of Chaotic Attractors, Springer-Verlag, Berlin

[44] Iwanik A (1989) Independent sets of transitive points in: Dynamical Systems and Ergodic Theory, Banach Centre Publ. 23:277-282

[45] Katok A and Hasselblatt B (1995) Introduction to the Modern Theory of Dynamical Systems, Cambridge Univ. Press, Cambridge

[46] King JF, Entropy, this encyclopedia

[47] Kuratowski K (1973) Applications of the Baire-category method to the problem of independent sets, Fundamenta Math. 81:65-72

[48] Li TY and Yorke J (1975) Period three implies chaos, Amer. Math. Monthly 82:985-992

[49] Lorenz E (1993) The Essence of Chaos, University of Washington Press, Seattle

[50] May RM (1973) Stability and Complexity in Model Ecosystems, Princeton Univ. Press, Princeton

[51] Murdock J (1991) Perturbations, John Wiley and Sons, New York.

[52] Mycielski J (1964) Independent sets in topological algebras, Fundamenta Math. 55: 139-147

[53] Nemytskii V and Stepanov V (1960) Qualitative Theory of Differential Equations, Princeton U Press, Princeton

[54] Oxtoby J (1980) Measure and Category (2nd ed), Springer-Verlag, Berlin

[55] Peterson K (1983) Ergodic Theory, Cambridge Univ. Press, Cambridge

[56] Robinson C (1995) Dynamical Systems: Stability, Symbolic Dynamics and Chaos, CRC Press, Boca Raton

[57] Rudolph D (1990) Fundamentals of Measurable Dynamics: Ergodic Theory on Lebesgue Spaces, Oxford Univ. Press, Oxford

[58] Shub M (1987) Global Stability of Dynamical Systems, Springer-Verlag, New York

[59] Shub M and Smale S (1972) Beyond hyperbolicity, Ann. of Math 96:587-591

[60] Sigmund K (1993) Games of Life: Explorations in Ecology, Evolution and Behavior, Oxford Univ. Press, Oxford

[61] Smale S (1967) Differentiable dynamical systems, Bull. Amer. Math. Soc. 73:747-817

[62] Smale S (1980) The Mathematics of Time, Springer-Verlag, Berlin

[63] Stewart I (1989) Does God Play Dice?, Basil Blackwell, Oxford

[64] Ura T (1953) Sur les courbes défines par les équations différentielles dans l'espace à m dimensions, Ann. Sci. Ecole Norm. Sup.[3] 70:287-360

[65] Ura T (1959) Sur le courant extérieur à une région invariante, Funk. Ehv. 2:143-200

[66] de Vries J (1993) Elements of Topological Dynamics, Kluwer, Dordecht, Holland