Conference Board of the Mathematical Sciences

CBMS

Regional Conference Series in Mathematics

Number 95

Single Orbit Dynamics

Benjamin Weiss

Published for the
Conference Board of the Mathematical Sciences
by the
American Mathematical Society
Providence, Rhode Island
with support from the
National Science Foundation

Contents

Preface

These notes represent a mildly expanded version of a series of ten lectures that I gave at a CBMS conference organized by Kamel Haddad at California State University, Bakersfield, CA in June, 1995. Due to external circumstances their publication has been delayed for a few years but I hope that they still give a timely presentation of a novel point of view in dynamical systems. I have not made any effort to update the exposition, but I would like to point out some recent developments that are relevant to the reader who wishes to pursue matters further.

First of all I would like to recommend the new book by Paul Shields, *The ergodic theory of discrete sample paths*, Grad. Studies in Math. v. 13 (AMS), 1996. In this book there is a very careful treatment of some central issues in ergodic theory and information from a point of view that is close to that expounded in these lectures. There is a particularly good treatment there of entropy related matters and of various characterizations of Bernoulli processes.

Following up on an idea proposed by M. Gromov, Elon Lindenstrauss and I have developed a new invariant in topological dynamics which refines the classical notion of topological entropy. This invariant, called the mean topological dimension, vanishes for all systems with finite topological entropy but distinguishes between various systems with infinite topological entropy. There is a single orbit interpretation of this invariant which should shed some new light on spaces of meromorphic functions, solutions of dynamical systems with infinitely many degrees of freedom etc. The basic theory is set out in a joint paper *Mean Topological Dimension*, (to appear in the Israel J. of Math).

The style of these notes is that of a lecture. When proofs are given they are meant to be complete, but not every i is dotted nor every t crossed. Most of the material appears elsewhere and I have given references at the end of each chapter to guide the reader who wants to pursue matters in more detail. Chapter 4 contains

CHAPTER 1

What is Single Orbit Dynamics

The study of dynamical systems has its origins in the classical mechanics of Newton and his successors. The modern theory has flowered in several different directions that are best described by abstracting certain features of the classical systems. The ergodic theorem of J. von Neumann pointed the way to the functional analytic **- operator ergodic theory** while the seminal ergodic theorem of G.D. Birkhoff founded what we now call **measure ergodic theory** which can be viewed as the study of transformations of measure spaces. Keeping only the topological structure led to the **topological dynamics** of Gottschalk and Hedlund and focusing on the analytic or smooth structures gave rise to **complex dynamics** and **smooth dynamics**. Each of the above has expanded to a broad and fairly independent discipline with its own problems and methodology. Less well known specialties of a similar nature are **generic dynamics** where first category, in the sense of Baire, replaces the μ-null sets as the negligible sets, and **measurable or Borel dynamics** where no sets are neglected a priori and the only structure that comes into play is the Borel structure.

Single orbit dynamics is not another kind of dynamics of the type described above. It is rather an attempt to focus attention on a rather large body of work concerned with dynamical study of single orbits as opposed to the global study of a system as a whole. One can see the origins of this point of view in the works of R. von Mises, H. Bohr and N. Wiener none of whom actually thought about what they were doing in the terms that I just used. Nonetheless, I will begin by telling something about the relevant work of each of these mathematicians before I try to give some general formulation of single orbit dynamics.

At the beginning of the century probability theory was not yet established as a bona fide member of the community of mathematics. Its status was closer to that of a physical theory where the phenomena that were being described were games of chance and other random events and processes. Richard von Mises attempted to base a mathematical theory of probability on the primitive notion of a **collective**. This was supposed to capture the idea of a random sequence of outcomes (H-heads or T-tails) of a simple experiment such as tossing a coin. His axiomatic description of such a sequence of simple coin tossing was basically:

I. The asymptotic frequency of occurrences of H in the collective equals $1/2$.

II. Property I persists for any subsequence of outcomes derived from the collective by place selection rule.

In order to avoid misrepresenting his ideas here is a direct quote from an English version of his book Probability, Statistics and Truth posthumously published by his widow Hilda Geiringer:

A collective appropriate for the application of the theory of probability must fulfill two conditions. First, the relative frequencies of the attributes must possess limiting values. Second, these limiting values must remain the same in all partial sequences which may be selected from the original one in an arbitrary way. Of course, only such partial sequences can be taken into consideration as can be extended indefinitely, in the same way as the original sequence itself. Examples of this kind are, for instance, the partial sequences formed by all odd members of the original sequence, or by all members for which the place number in the sequence is the square of an integer, or a prime number, or a number selected according to some other rule, whatever it may be. The only essential condition is that the question whether or not a certain member of the original sequence belongs to the selected partial sequence should be settled **independently of the result** *of the corresponding observation, i.e., before anything is known about this result. We shall call a selection of this kind a* **place selection**. *The limiting values of the relative frequencies in a collective must be independent of all possible place selections. By place selection we mean the selection of a partial sequence in such a way that* *we decide whether an element should or should not be included without making use of the attribute of the element, i.e., the result of our game of chance.*

While this may appear strange to generations of mathematicians who have been raised on Kolmogorov's treatment of probability theory as a special kind of measure theory it represented an earlier, systematic mathematical foundation of probability. What is important for us is his insistence that one can calculate probabilities from a **single** sequence of outcomes.

During the 1930's the approach of von Mises fell into disfavor, after the criticisms of Tornier, Ville, Frechet and others who pointed out the difficulties in establishing the existence of collectives. The criticisms hinge, of course, on what interpretation one places on the phrase "admissible place selections". We shall discuss these matters in more detail in a later lecture after we will have some ergodic theoretic tools at our disposal.

Next we turn to H. Bohr's theory of almost periodic functions. Motivated by his attempts to prove the famous Riemann hypotheses on the zeros of the zeta function he was led to consider bounded functions on the reals $f(t)$ that generalized the usual periodic functions. Here is his definition:

DEFINITION 1.1. A complex valued bounded function on $\mathbb{R}$, $f(t)$ is **almost periodic** if for any $\varepsilon > 0$, the numbers $p \in \mathbb{Z}$ that satisfy

$$\sup_n |f(t) - f(t+p)| \leq \epsilon \tag{1}$$

are **syndetic**, i.e. the gaps between successive p's are bounded.

Any p which satisfies (1) is called an almost period. Bohr showed that just like periodic functions can be expanded in a Fourier series so too the almost periodic functions have an expansion as a generalized Fourier series. More precisely we have:

THEOREM 1.2. *If f is almost periodic then there is a countable set $\Lambda \subset \mathbb{R}$ such that for all $\epsilon > 0$ there is an exponential polynomial*

$$Q(t) = \sum_{j=1}^{J} c_i e^{i\lambda_j t}, \qquad \{\lambda_1, \dots, \lambda_J\} \subset \Lambda$$

that satisfies

$$\sup_t |Q(t) - f(t)| \leq \epsilon.$$

This class of functions and many of its generalizations became an active field of investigation and turned out to have many applications. We mention one later result due to Favard:

THEOREM 1.3. *If $f(t)$ is a bounded function such that*

$$f(t+1) - f(t) = g(t)$$

is almost periodic then $f(t)$ is also an almost periodic function.

On the face of it there seems to be no connection between this set of ideas and dynamical systems. Nonetheless, S. Bochner pioneered an approach to the theory of almost periodic functions which fits perfectly into the framework of topological dynamics and in particular to what we mean by single orbit dynamics. This will be explained in detail in the next lecture.

Our third historical example is the generalized harmonic analysis of N. Wiener. Motivated by problems in the newly emerging field of electrical engineering and communication theory he was led to define the **spectrum** of a function $\xi(n)$ as the set of complex numbers σ of modulus one such that

$$\lim_{N\to\infty} \frac{1}{2N+1} \sum_{-N}^{N} \bar{\sigma}^n \xi(n) \neq 0.$$

Naturally for sequences to have a spectrum it is necessary that this limit exist for all σ with $|\sigma| = 1$. As he later showed, stationary stochastic processes, provide a rich source for such sequences. We shall discuss these matters in the third lecture, for now it suffices for us to see here another example of some kind of nontrivial analysis of a single sequence being the center of a series of investigations.

All three of these examples, the collectives of von Mises, the almost periodic functions of Bohr and the time series of Wiener illustrate how individual functions are important objects of investigations. As we will see, all three of these examples can be very profitably studied using certain concepts and techniques drawn from dynamical systems. This is one aspect of single orbit dynamics. Here are some more recent examples of this type that center around the general theme of patterns in large sets. The most famous of this class of results is the theorem of van der

Waerden that in any finite partition of the integers at least one of the sets contains arithmetic progressions of arbitrary length.

About twenty years ago, H. Furstenberg and his collaborators showed how ideas from dynamical systems could be used to obtain old and new results of this type. The basic paradigm involved interpreting the object under investigation as a single orbit of an entire dynamical system. This study of patterns in large sets has developed into a broad discipline and has injected powerful new methods into basic combinatorial questions. I shall describe in a little more detail, now, a more recent example of this which we shall discuss in detail in Chapter 5.

For any set D of positive integers we define a graph $G(D)$ with vertex set $\mathbb{Z}$ and edges consisting of all pairs (i, j) such that $|i - j|$ belongs to D. This $G(D)$ is a translation invariant graph on the integers. A very natural question is: when does $G(D)$ have a finite chromatic number? This question can be translated into a question of topological dynamics and then some standard methods there lead, for example, to the result that whenever D is **lacunary** $G(D)$ has a finite chromatic number. On the other hand, for any growth rate that is less than lacunary there are sequences D with that growth rate for which the chromatic number of $G(D)$ is infinite.

In all of the above we are dealing with the **existence** of patterns. It turns out that the **rate** of recurrence of patterns can also be examined using dynamical tools. If $x = x_1x_2x_3\ldots$ is an infinite sequence on a finite alphabet A, we can define

$$R_n(x) = \min\{r > n : x_{r+i} = x_i \qquad 1 \le i \le n\}$$

if there is such an r, and set $R_n(x) = \infty$ otherwise. For regular sequences x that are generated by stationary stochastic processes it turns out that $R_n(x)$ has a well defined growth rate that is intimately related to Shannon's entropy. This connection makes it possible to give sense to the information content of individual sequences.

Up to now the emphasis has been on that aspect of single orbit dynamics in which the main object of interest is a single orbit and its properties and the global dynamics is a tool. The reverse situation, in which the main object of interest is the global system and the individual orbit is but a tool is also not uncommon. We call this the principle of **one for all**. In topological dynamics this means describing a global system as the orbit closure of a single orbit. This is a powerful tool in the construction of a variety of examples. A classical example is the Morse minimal sequence defined as follow: our alphabet will be $\{0,1\}$, and we agree that $\bar{0}=1$, $\bar{1}=0$. If B is a block consisting of k-symbols then $\bar{B}$ is the k-block obtained by replacing each $b \in B$ by $\bar{b}$. Define inductively:

$$\begin{aligned} &B_0 = 0, \\ &\ \vdots \\ &B_{n+1} = B_n \bar{B}_n \\ &\ \vdots \end{aligned}$$

then $B_n \to x = x_0 x_1 x_2 \ldots$ which is called the Morse minimal sequence. It begins as follows

$$x = 0110100110010110\ldots$$

and its orbit closure in $\{0,1\}^{\mathbb{N}}$ under the shift map provides a very basic example in topological dynamics.

In ergodic theory orbit closures alone are not sufficient and one adds the notion of a **generic point** which allows one to define probabilities in terms of a single orbit. Statistical investigations in which one tries to learn the global features of a system from an individual sample sequence also are covered by this principle of **one for all** in the probabilistic framework. We will study several examples of this in greater detail below. These include questions like the classical ones:

- How do the successive observations $\xi_1 \xi_2 \xi_3 \ldots$ of a stationary stochastic process enable one to give a better and better description of the process (as n tends to infinity)?
- How well can one predict the next output ξ_1 as one learns more and more about the past $\xi_0 \xi_{-1} \xi_{-2} \ldots$?

At this point I would like to give a concrete example of this single orbit philosophy in action. A very classical result in probability theory is the recurrence of the one dimensional simple random walk. Let $\{x_n\}_1^\infty$ be independent random variables with distribution

$$P(x_n = +1) = P(x_n = -1) = 1/2 \qquad \text{for all } n,$$

and define

$$S_n = x_1 + x_2 + \cdots + x_n.$$

The recurrence of this simple random walk is the statement that with probability one S_n will take the value 0. It is known that the same holds true if the increments x_n come from any stationary stochastic process with zero mean. Not so well known is the following generalization.

THEOREM 1.4. *Let $\{x_n\}$ be a $\mathbb{Z}$-valued stationary stochastic process such that for all $\epsilon > 0$*

$$P\left\{ \left| \frac{1}{n} \sum_1^n x_i \right| > \epsilon \right\} \to 0 \qquad \textit{as} \;\; n \to \infty.$$

Then with probability one $S_n = 0$.

Even for independent random variables this is a stronger result than the usual one since the hypothesis may be satisfied even when $E\{|x_i|\}$ is infinite.

Denote by Ω the underlying probability space of the random variables x_i, and fix a small $\epsilon > 0$. Let

$$E_n = \left\{ \omega : \quad \left| \frac{1}{n} \sum_1^n x_i(\omega) \right| \le \epsilon^2 \right\}$$

and let N_0 be sufficiently large so that for all n larger than N_0, $P(E_n) > 1 - \frac{\epsilon^2}{2}$, and then let N be much larger than N_0 $\bigl(\text{say } N > \left(\frac{2}{\epsilon^2} N_0\right)\bigr)$. Consider now

$$f(\omega) = \frac{1}{N} \sum_1^N 1_{E_n}(\omega).$$

We have arranged matters so that the integral of f is greater than $1 - \epsilon^2$, and since $0 \le f \le 1$ it follows that on most of the space $f(\omega) \ge 1 - \epsilon$.

We focus now on a single ω for which $f(\omega) \ge 1 - \epsilon$ and consider the graph of

$$S_n(\omega) = \sum_1^n x_i(\omega)$$

thought of as a function from $\{1, 2, \ldots, N\}$ to $\mathbb{Z}$. For most of the n's we have that $|S_n(\omega)| \le \epsilon N$, and thus only a small fraction of the indices n_0 can have the property that the value $S_{n_0}(\omega)$ is not repeated for some $n_1 > n_0$. Thus if we view the progress of the random walk along this typical ω, it turns out that for most of our visits $S_{n_0}(\omega)$ that site is visited again at a later time $n_1 \leq N$.

Up to this point we haven't used the stationarity of the increment process. Now this stationarity will be used to translate the recurrence at later times to recurrence to the origin - 0 - our position when the walk begins. We again use an averaging argument. For a set with probability at least $1 - \epsilon$ we know that $1 - 2\epsilon$ of the indices $n_0 \leq N$ have the property that there is an $n_0 < n_1 \leq N$ for which

$$S_{n_1}(\omega) = S_{n_0}(\omega).$$

It follows that for some **fixed** n_0, there is a set of ω of probability at least $1 - 3\epsilon$ so that for each of these ω's there is an $n_1 > n_0$ with $S_{n_1}(\omega) - S_{n_0}(\omega) = 0$. What this means is that if we would define a random walk with increments $x_{n_0+1}, x_{n_0+2}, \dots$ then this one would return to the origin with probability at least $1 - 3\epsilon$. But by stationarity this is the same as the probability that the original random walk returns to the origin. Since ϵ was arbitrary this completes the proof of the theorem. □

The proof of the theorem takes place by analyzing closely the behavior of a single orbit $S_n(\omega)$, with "Fubini's theorem", or "averaging in different ways" being used to translate the global hypothesis, the weak law of large numbers for the x_n's, into information concerning a single orbit. This type of interplay between the individual orbits and the global properties will be a recurring theme in the rest of these lectures.

The final point that I would like to make in this introductory lecture is the distinction between almost everywhere results and single orbit theorems. I will illustrate this with two examples - one drawn from von Mises collectives and the other from Wiener's generalized harmonic analysis.

DEFINITION 1.5. A number $x \in (0, 1)$ is normal in the base b if in the b-ary expansion of x

$$x = \frac{\xi_1}{b} + \frac{\xi_1}{b^2} + \cdots + \frac{\xi_i}{b^i} + \cdots \xi_i \in \{0, 1, \dots, b-1\}$$

any block $B \in \{0, 1, \dots, b-1\}^K$ occurs with an asymptotic frequency equal to b^{-K}, $(K = 1, 2, \dots)$.

More formally the condition means that for all such B;

$$\lim_{n\to\infty} \frac{1}{n}|\{i \le n : \ \xi_i\xi_{i+1}\cdots\xi_{i+k-1} = B\}| = b^{-K}.$$

At the beginning of the century, E. Borel proved that the set of non-normal numbers has Lebesgue measure zero. It is an easy consequence of his proof that for any sequence $n_1 < n_2 < \dots$ a.e. x has the property that $\xi_{n_1}\xi_{n_2}\xi_{n_3}\dots$ is also a normal number. These are examples of a.e. theorems. The following theorem characterizes those positive density sequences $\{n_j\}$ that have the property that $\xi_{n_1}\xi_{n_2}\dots$ is normal whenever x is normal. This is a single orbit theorem since it asserts something about individual orbits and conclusions can be drawn about specific points. For the formulation of the theorem a preliminary definition is necessary.

Let us denote by Ω the space $\{0,1\}^{\mathbb{N}}$ with the product topology and let $T : \Omega \to \Omega$ denote the shift on Ω, so that $(T\omega)(n) = \omega(n+1)$. Probability measures on Ω that are T-invariant define $\{0,1\}$-valued stationary stochastic processes (the coordinates $x_n(\omega) = \omega(n)$) and vice versa. Such a process is called **deterministic** if x_0 is measurable with respect to the σ-field generated by the $\{x_i : \ i \ge 1\}$. The topology on the space of T-invariant measures that we consider is the w^*-topology which is such that $\mu_n \to \mu$ if for all finite cylinder sets C $\mu_n(C)$ converges to $\mu(C)$.

Definition 1.6. A sequence $u \in \{0,1\}^{\mathbb{N}}$ is said to be **completely deterministic** if any invariant probability measures that is in the closure of the set of measures:

$$\nu_N = \frac{1}{N}\sum_{n=1}^{N} \delta_{T_u^n}$$

defines a deterministic process.

Theorem 1.7 (T. Kamae - B. Weiss). *A positive density sequence $\{n_1 < n_2 < \dots\}$ has the property that for every normal number*

$$x = .\xi_1\xi_2\xi_3\dots$$

also $.\xi_{n_1}\xi_{n_2}\xi_{n_3}\dots$ is normal if and only if the indicator function of the sequence is completely deterministic.

The indicator function of the sequence $\{n_j\}$ is the element $u \in \{0,1\}^{\mathbb{N}}$ such that $u(n) = 1$ if and only if n is one of the n_j's. Simple examples of completely deterministic sequences are $[j\alpha]$ for any real $\alpha \geq 1$. Note that there are uncountably many distinct completely deterministic sequences.

A generalization of the notion of a normal number is the following. If T is a continuous mapping of a compact metric space X to itself and μ is a T-invariant probability measure on X then $x_0 \in X$ is said to be generic for the system (X, T, μ) if for all continuous functions f on X we have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} f(T^i x_0) = \int_X f(x) d\mu(x).$$

For the coordinate shift on Ω above, being a normal number coincides with being a generic point for the product measure

$$\prod_0^\infty \left(\frac{1}{2}, \frac{1}{2}\right) \quad \text{on} \quad \Omega$$

with respective to T, the coordinate shift.

Here is a result of Wiener-Winter that applies to **all** generic points of a system.

THEOREM 1.8. *If x_0 is generic for (X, T, μ) and the only eigenfunctions that T has as an operator on $L^2(X, \mu)$ are the constants then for all complex λ of modulus one $\lambda \neq 1$ and all continuous functions*

$$\lim_{N\to\infty} \frac{1}{N} \sum_1^N \lambda^j f(T^j x_0) = 0.$$

This is the single orbit result from the generalized harmonic analysis of N. Wiener. It provides a rich class of examples of bounded sequences for which a spectral analysis is possible. One way to use such a result is to observe that the failure of the limit above to vanish for some λ means that the system has some point spectrum.

To sum up briefly, we have discussed several kinds of results involving single orbits — ranging from the analysis of individual sequences through the construction of entire systems from single orbits up to conclusions that are valid for all generic points of certain systems. These results all point to the centrality of the behavior of individual orbits and to highlight this focus we call this body of work single orbit dynamics. In the following lectures we will illustrate all of these themes by concrete examples.

References

1. G.D. Birkhoff, *Dynamical systems*, AMS Colloq. 9, Providence, R.I., 1927.
2. H. Bohr, *Almost periodic functions*, Chelsea, New York, 1951.
3. J. Favard, *Sur les equations differentielles lineaires à coefficients presque periodiques*, Acta Math. **51**(1928), 31–81.
4. H. Furstenberg, *Stationary processes and prediction theory*, Annals of Math Studies 44, Princeton, N.J., 1960.
5. H. Furstenberg, *Recurrence in Ergodic Theory and Combinatorial Number Theory*, Princeton, N.J., 1981.
6. W.H. Gottschalk and G.A. Hedlund, *Topological Dynamics*, AMS Colloq. 36, Providence, R.I., 1955.
7. T. Kamae, *Subsequences of normal sequences*, Israel J. of Math. **16**(1973), 121–149.
8. Richard von Mises, *Probability Statistics and Truth*, revised English edition by Hilda Geiringer, London, NY, 1957.
9. B. Weiss, Normal Sequences as Collectives, in *Proceeding of Symposium on Topological Dynamics and Ergodic Theory*, Univ. of Kentucky, 1971.
10. B. Weiss and T. Kamae, *Normal Numbers and selection rules*, Israel J. of Math. **21**(1975), 101–110; Israel J. of Math. **21**(1975), 159–166.
11. B. Weiss, *Measurable Dynamics*, in Conference in Modern Analysis and Probability, Cont. Math. **25**(1984), 395–421.
12. B. Weiss, D. Sullivan and J. Wright, *Generic dynamics and monotone complete C^*-algebras*, TAMS, **295**(1986), 795–809.
13. B. Weiss and D. Ornstein, *How sampling reveals a process*, Annals of Prob. **18**(1990), 905–930.
14. B. Weiss, H. Furstenberg and Y. Katznelson, *Ergodic theory and configurations in sets of positive density*, in Mathematics of Ramsey Theory, ed. by J. Nesetril and V. Rodl, Springer, 1990, pp. 184–199.
15. N. Wiener, *Generalized harmonic analysis*, Acta Math. **55**(1930), 117–258.

results that haven't appeared before in print. They originate in discussions that I had with Don Ornstein fifteen years ago when we were traveling on a weekly basis from Stanford to MSRI. The main result in chapter 5 is due to Y. Katznelson, I thank him for his permission to include it in these lectures. Hillel Furstenberg's influence on these lectures began with a course that I took with him thirty five years ago at Princeton during which he gave the first exposition of his ideas on disjointness which he then called absolutely independent. It has continued ever since and culminated in a careful reading that he, and Eli Glasner, gave of these notes. Naturally the responsibility for all remaining errors is mine alone.

Finally I would like to express my thanks to Kamel Haddad who organized the wonderful conference that made these notes possible. Last, but not least my thanks to Stanford University and Willene Perez, who typed there the first draft of these notes, and to Shani Ben David, the Hebrew University of Jerusalem, who is responsible for the final version.

Benjamin Weiss, July 1999

Jerusalem (TVBBA), Israel