

ON THE GENESIS OF SYMBOLIC DYNAMICS
AS WE KNOW IT

BY

ETHAN M. COVEN (Middletown, CT) and
ZBIGNIEW H. NITECKI (Medford, MA)

Abstract. We trace the beginning of symbolic dynamics—the study of the shift dynamical system—as it arose from the use of coding to study recurrence and transitivity of geodesics. It is our assertion that neither Hadamard’s 1898 paper, nor the Morse–Hedlund papers of 1938 and 1940, which are normally cited as the first instances of symbolic dynamics, truly present the abstract point of view associated with the subject today. Based in part on the evidence of a 1941 letter from Hedlund to Morse, we place the beginning of symbolic dynamics in a paper published by Hedlund in 1944.

Symbolic dynamics, in the modern view [LM95, Kit98], is the dynamical study of the shift automorphism on the space of bi-infinite sequences of symbols, or its restriction to closed invariant subsets. In this note, we attempt to trace the beginnings of this viewpoint. While various schemes for symbolic coding of geometric and dynamic phenomena have been around at least since Hadamard (or Gauss: see [KU07]), and the two papers by Morse and Hedlund entitled “Symbolic dynamics” [MH38, MH40] are often cited as the beginnings of the subject, it is our view that the specific, abstract version of symbolic dynamics familiar to us today really began with a paper, “Sturmian minimal sets” [Hed44], published by Hedlund a few years later. The outlines of the story are familiar, and involve the study of geodesic flows on surfaces, specifically their recurrence and transitivity properties; this note takes as its focus a letter from Hedlund to Morse, written between their joint papers and Hedlund’s, in which his intention to turn the subject into a part of topology is explicit ⁽¹⁾. This letter ⁽²⁾ is reproduced on pages 235–236, along with Morse’s reply on page 238.

2000 *Mathematics Subject Classification*: Primary 37B10; Secondary 01A60.

Key words and phrases: symbolic dynamics, shift automorphism, recurrence, transitivity, minimal set, geodesics.

⁽¹⁾ In the interest of full disclosure, we should indicate our possible bias: Hedlund was the first author’s dissertation director and hired the second author in his first job.

⁽²⁾ More accurately, a carbon copy of the original provided to the first author by Hedlund.

Our focus here is rather narrow, even when it comes to coding geodesic flows. A recent survey by Katok and Ugarcovici [KU07] distinguishes two approaches to such coding: a geometric method, which is our subject, and a second, going back to Gauss and associated with Artin, Koebe and Nielsen, which can be regarded as more arithmetic in nature. We do not propose to consider the latter in detail. The survey [KU07] discusses technical details of both approaches as well as their subsequent development in recent years.

We have focused here on the line of literature from Hadamard to the Morse–Hedlund papers, and have not attempted to relate this to the concurrent development of ergodic theory, in which of course Birkhoff (who plays a major role in our story) was very much involved. In contrast to the present use of symbolic methods in ergodic theory, there seems to have been little if any such connection drawn in this earlier period: for example, Hopf’s influential book [Hop37] shows no signs of symbolic coding. Similarly, we have not attempted to explore the contribution of the Russian school to this development: in his 1949 survey of dynamical systems, Nemytskiĭ explicitly credits Morse with the idea of using symbolic coding for the trajectories of a system [Nem49, p. 20 of English version] ⁽³⁾.

The beginnings of symbolic dynamics are often traced back to Hadamard’s 1898 study of geodesics on surfaces of negative curvature [Had98]. In Part II of that work (§20), Hadamard gives a coding for the (free homotopy classes of) closed geodesics, essentially as words (up to cyclic permutation) in generators of the fundamental group, and in Part III (§37) he shows that each word corresponds to a unique closed geodesic. He then goes on to study unbounded geodesics and in Part VI (§56) shows that the initial conditions at a point which determine geodesics staying in a bounded region form a perfect, nowhere dense closed set (which is the closure of the conditions yielding closed geodesics).

There are several respects in which Hadamard’s paper does not really qualify as a beginning for symbolic dynamics. First, his coding is limited to *finite* words, coding *closed* geodesics; he does not appear to envision a coding system encompassing other geodesics. The bounded, non-periodic geodesics he produces in Part VI are, in passing, seen as determined by a sequence of closed geodesics, but this is not explicitly related to their coding. Furthermore, Part II of the paper (where the coding is formulated), entitled “Considérations d’Analysis situs”, is presented as follows (in a rough

⁽³⁾ In fact, Nemytskiĭ describes a tantalizing example from an unpublished manuscript by Pontryagin and Shnirel’man that is essentially the shift automorphism; unfortunately, there is neither a date nor any further bibliographic information provided for this work, and so it is difficult to assess its historic importance.

translation ⁽⁴⁾):

Having established, in the preceding sections, the existence of surfaces of negative curvature and of arbitrary connectivity ⁽⁵⁾, we will recall the principles which govern the study of curves on these surfaces, principles set forth by Jordan in a well-known Memoir (2).

Footnote (2) (“Ce journal, année 1866”) refers to the second of two back-to-back papers published 33 years before Hadamard in the *Journal de Mathématiques Pures et Appliquées* by Jordan [Jor66b, Jor66a]: the first concerns the role of fundamental contours in determining the homeomorphism type of a surface, and the second presents the notion of a “class” of contours (i.e., free homotopy class) subsequently used by Hadamard in [Had98]. Jordan’s notation is that adopted by Hadamard, and he hints at the representation of curves by words (with positive or negative exponents). Finally, it should be noted that in Hadamard’s study, the point of view is geometric rather than dynamic: geodesics are regarded as oriented curves, and there appears no explicit sense of a “geodesic flow”; in particular Hadamard’s symbolic coding is static in nature.

In an important paper [Bir12] published in 1912 (and based on a presentation to the American Mathematical Society in 1909), G. D. Birkhoff analyzes the behavior of recurrent trajectories in a dynamical system defined by a system of ordinary differential equations. The word “recurrent” here corresponds to what we now call “minimal” ⁽⁶⁾. A collection M of trajectories of a dynamical system is *minimal* if every element of M has all elements of M in its α - and ω -limit sets; Birkhoff calls any trajectory belonging to a minimal set “recurrent”. He proves that recurrence (in this sense) is equivalent to (what we now call) *almost-periodicity*: for any $\varepsilon > 0$ there exists a length T such that the whole trajectory is contained in an ε -neighborhood of any segment of length T . Obvious examples of minimal sets are equilibria and closed orbits; Birkhoff also notes the example of dense lines on a torus, and calls a recurrent motion *continuous* if the corresponding minimal set forms a continuum of some dimension. The *a priori* possibility

⁽⁴⁾ “Ayant reconnu, dans les numéros précédents, l’existence de surfaces à courbures opposées et à connexion quelconque, nous avons à rappeler les principes qui gouvernent l’étude des lignes tracées sur de telles surfaces : principes posés par M. Jordan dans un Mémoire bien connu (2).”

⁽⁵⁾ The connectivity of a surface is the maximal number of non-separating curves; for a closed surface it equals the genus plus one [Hir76, pp. 188–189]. Of course, there are topological restrictions on which surfaces can carry a complete metric of negative curvature, but this does not restrict the connectivity if the surface is allowed to be non-compact.

⁽⁶⁾ What we now call recurrent is called *stable in the sense of Poisson*.

of *discontinuous* recurrent trajectories is illustrated by the suspension of a nontransitive homeomorphism of the circle with irrational rotation number, and Birkhoff asks whether discontinuous recurrent trajectories can occur in analytical dynamical systems.

Morse, in his 1917 dissertation under Birkhoff (published as [Mor21a] and [Mor21b]) establishes the existence of recurrent geodesics of discontinuous type on surfaces of negative curvature and negative Euler characteristic. He considers the bounded region S obtained by cutting off any infinite “funnels” using closed geodesics, and codes the geodesics entirely contained in S by recording the order in which they cross a family of transversals (“normal segments”) that cut S into a simply-connected region—in effect lifting the geodesic to the hyperbolic plane. He then shows that this coding distinguishes geodesics in S , and by constructing the “Morse sequence” (discovered earlier and independently by Thue [Thu12]) proves the existence of discontinuous recurrent geodesics. Furthermore, he shows by symbolic methods that every closed geodesic in S is a limit of discontinuous recurrent ones. Since Hadamard had shown that every geodesic in S is a limit of closed geodesics, it follows that the recurrent geodesics of discontinuous type are dense in the set of all geodesics contained in S . Despite the closer connection with dynamical ideas, the point of view in these papers remains geometric: geodesics are still regarded as curves rather than trajectories, and the coding is used to establish that a geodesic is recurrent (in his sense) and not closed.

In 1920, Birkhoff published a study [Bir20] setting forth a number of ways that the behavior of a dynamical system with two degrees of freedom can be studied by means of the successive intersections of orbits with a transverse surface; strictly speaking, such a *surface of section* is not entirely transverse to the flow, as it is bounded by closed orbits, but its interior *is* transverse to the flow. The general setup had been formulated by Poincaré in [Poi97, Vol. III, Chap. 27], as a means of studying periodic and homoclinic orbits in celestial mechanics. Birkhoff had used the same setup in a limited way for similar purposes in 1917 ([Bir17]). In Chapters 5 and 6 of [Bir20], Birkhoff goes beyond the study of periodic orbits to study, in some abstraction, the behavior of the “Poincaré map”, defining α - and ω -limit sets, minimal sets and his related notion of recurrence, as well as transitivity.

In 1924, Artin published a brief but influential paper [Art24] in which he shows that the orbit space of the group of linear fractional transformations with integer coefficients acting on the hyperbolic plane (in the half-plane model) has a dense geodesic (in fact, the set of these has full measure). His proof involves a coding of geodesics via the continued fraction expansion of their “endpoints at infinity” on the real line. (His term for transitivity is “quasiergodicity”.)

In 1927, in the first of his papers on mapping classes [Nie27], Nielsen formulates a similar coding geometrically, in terms of the fundamental group, to study the axes of hyperbolic transformations on surfaces obtained as quotients of the hyperbolic disc by a Fuchsian group. Nielsen's approach has some similarities to Morse's coding of geodesics via transverse segments, but the dynamics that comes in is that of the Fuchsian group acting on the universal covering.

In his 1927 book, *Dynamical Systems* [Bir27], a broad survey of work on dynamical systems (primarily of mechanical origin), Birkhoff included Chapter 7, "General theory of dynamical systems", which sets forth the notions of wandering and non-wandering orbits, central motions, minimal sets, and transitivity in the general context of the flow generated by a system of differential equations. Much of this reflected ideas formulated earlier in his 1912 paper [Bir12].

In 1935, Birkhoff summed up his work on dynamics in a long paper [Bir35], published by the Pontifical Academy of Sciences. Chapter 3, a study of behavior near a hyperbolic periodic orbit, is based on a detailed examination of the dynamics of a Poincaré map for a transverse section. By symbolic methods that, several decades later, were modified and used by Smale to prove the "Birkhoff–Smale" theorem ⁽⁷⁾ and to construct the "horseshoe", Birkhoff demonstrates the existence of highly complicated "first return" behavior for periodic orbits near any orbit homoclinic to a hyperbolic periodic orbit or, more generally, belonging to a loop of heteroclinic connections.

This work forms the background to two papers by Morse and Hedlund, entitled "Symbolic dynamics" [MH38, MH40], published in 1938 and 1940, respectively.

Hedlund, in his dissertation written under Morse in 1929 [Hed32a, Hed32b] had proved the existence of a length-minimizing closed geodesic in each free homotopy class for any Riemannian metric on the torus; Morse [Mor24] had proved the same result for surfaces of higher genus. Hedlund went on to study geodesics on surfaces [Hed35b, Hed35a, Hed36b, Hed36a], in particular proving the ergodicity of the geodesic flow on a closed surface of constant negative curvature [Hed35b], using Nielsen's symbolic coding [Nie27], and transitivity of the horocycle flow [Hed36a]. In 1939 he published a survey of results on the dynamics of geodesic flows [Hed39], in which he formulates seven types of transitivity, elaborating on Birkhoff's definitions in [Bir27]: these include topological transitivity ("regional transitivity", which he notes is equivalent to the existence of a dense trajectory), topological mixing ("permanent regional transitivity"), ergodicity with respect to some

⁽⁷⁾ That every neighborhood of a transverse homoclinic point contains a "horseshoe" [Sma65].

invariant measure (“metric transitivity”) and mixing (“mixture”) as well as hybrids of topological and ergodic notions of transitivity. He quotes theorems establishing many of these properties for geodesic flows on surfaces of constant negative curvature, as well as an example of a topologically mixing but non-ergodic geodesic flow. At the end of the article he plugs the work he had started with Morse in [MH38]:

The development of a symbolic theory apart from its dynamical significance has recently been begun by Morse and the author (cf. Morse [4]). This initial work includes an extensive analysis of transitive symbolic trajectories. The full scope of these symbolic methods in dynamics is yet to be determined.

The first of the Morse–Hedlund papers [MH38] sets forth a general theory of what we now call shift spaces, focusing on recurrence and transitivity properties of sequences. The motivation in the introduction refers primarily to geodesic flows on surfaces of negative curvature, but after that the treatment is quite abstract. The authors’ view of the place of their study in dynamics as a whole is stated as follows [MH38, pp. 816–817]:

Symbolic dynamics as the authors conceive it forms one of the three divisions

- (1) representation theory,
- (2) symbolic dynamics,
- (3) existence of space forms,

of the whole theory. The representation theory is concerned with the conditions on space forms under which trajectories admit a one-to-one symbolic representation in terms of which the recurrence or transitivity of the trajectory can be determined. These conditions will involve the Poincaré fundamental group of the space and differential conditions such as that of uniform instability (cf. Morse [4] ⁽⁸⁾, p. 64). In (3) one is concerned with the existence of space forms satisfying the conditions discovered in (1). The questions involved are rather deep extensions of the Hilbert, Koebe theory of spaces of negative curvature (cf. Hilbert [1] ⁽⁹⁾, and Koebe [1] ⁽¹⁰⁾). A simple typical theorem is that there exists no two-dimensional Riemannian manifold of the topological type of the torus satisfying the condition of uniform geodesic instability. The bearing of such studies on questions of

⁽⁸⁾ This reference is our [Mor21a].

⁽⁹⁾ [Hil01].

⁽¹⁰⁾ [Koe31].

topological and metric transitivity will be made clear in later papers.

Clearly, Morse and Hedlund view their paper as initiating a new branch of the theory of dynamical systems. However, it does not seem to us that the *shift dynamical system* is as yet considered as an object of study.

Beginning with a finite alphabet, Morse and Hedlund define an *I-trajectory* to be a two-sided indexed sequence of letters; a *symbolic element* $E(r, a)$ is an *I-trajectory* $a = \dots a_{-1}a_0a_1\dots$ together with a choice of a distinguished position r on it. The space of all symbolic elements is given the metric

$$d(E(r, a), E(s, b)) = 1/m$$

when $a_{r-m}\dots a_{r+m}$ and $b_{s-m}\dots b_{s+m}$ are the longest symmetric words centered on the distinguished positions which agree termwise (elements whose distinguished positions have different values are at infinite distance). They establish that this gives the space of symbolic elements the topology of a Cantor set. The space of *I-trajectories* is given the metric

$$[a, b] = \limsup_{n \rightarrow \infty} \frac{1}{2n+1} \sum_{i=-n}^n \delta(a_i, b_i)$$

(where $\delta(a_i, b_i) = 1$ or 0 according as a_i and b_i are the same or different), which they view as an analogue of the sup metric on functions, as used by Besicovitch in his treatment of almost-periodic functions [Bes32]—in fact, their notation closely follows his ⁽¹¹⁾. They define an *I-trajectory* a to be *almost-periodic* if for every $\varepsilon > 0$ the iterates D_r of the shift automorphism which satisfy

$$[D_r(a), a] < \varepsilon$$

form a relatively dense set of integers (that is, there is an integer $N = N(a)$ such that any set of N consecutive integers intersects the set). They consider subsets of the space of trajectories defined by admissible blocks; their admissibility rules appear to be of finite type, although they state a family of conditions [MH38, p. 823] which are far more restrictive, and appear to be motivated by Nielsen's formalism, to which they explicitly refer as an example. Again, they show that the subspace so defined has the topology of a Cantor set. They then study limit trajectories and minimal sets of trajectories from a symbolic point of view, and present the Morse sequence. The last 60% of the paper (pp. 833–864) is taken up with a number of functions that measure the “speed” of recurrence; these need not occupy us in detail here. It should be noted that, despite the dynamical background,

⁽¹¹⁾ Besicovitch's treatment, which is the one cited by Morse and Hedlund, was explicitly a generalization of earlier work by Bohr and Bochner.

and the appearance of the shift automorphism in two places (pp. 817 and 822), it is used in a way analogous to Besicovitch's use of translations to study almost-periodic functions (in fact, as we have noted, their notation is the same); there is no sense of a dynamical system generated by iteration of the shift.

Morse and Hedlund's second paper [MH40] concerns a specific class of subshifts, which they explain characterize the geodesics on a flat torus. These are built on an alphabet of two symbols and are defined by the condition that for each symbol, any two maximal blocks of consecutive appearances of the symbol differ in length by at most one. At the end of their previous paper, they had noted the relation of this condition to the Sturm Separation Theorem concerning the distribution of zeros of any non-constant solution of the differential equation

$$y'' + f(x)y = 0$$

where $f(x)$ is periodic with period one: one symbol represents the locations of zeros, the other the locations of integers (it is assumed without loss of generality that the solution has no integer zeros). They call such trajectories *Sturmian trajectories*. This paper is a detailed algebraic study of various combinatorial functions that characterize a Sturmian trajectory. Again, there is no explicit dynamical system here.

The journal lists [MH40] as received June 19, 1939. Two years later, Hedlund wrote to Morse (Figures 1 and 2).

Several features of this letter deserve comment. The first thing that strikes one is the penultimate paragraph, in which Hedlund wonders whether by a different choice of metric the space of symbolic elements could be made connected. Of course, on closer reading this is modified by the comment that this might be accomplished by identifying a countable number of pairs of points. Anachronistically, one could view this as a precursor of the construction of Markov partitions [Par66, Ber67, AW67, AW70]. Of course, the fact that the shift space itself has the topology of a Cantor set is these days taken for granted, and not viewed as problematic: the coding is a map from an initially given dynamical system to the shift dynamical system. Second, Hedlund is fully aware of the idea of a discrete dynamical system, and sees it as fundamentally equivalent to the idea of a flow. He is also completely aware of the interplay of topological and dynamical features of the system, something that is not clear in the earlier joint papers. Finally, the implied distinction between dynamics and topology has been somewhat erased in more recent years.

Perhaps a comment on the interests of "a number of topologists" mentioned in the first sentence of Hedlund's letter is in order. A search of the *Mathematical Reviews* (which began only a year before Hedlund's letter, in

Charlottesville

Virginia

June 7, 1941

Dear Marston:

As you probably know, a number of topologists are becoming interested in a study of the structure of the orbits obtained when a topological transformation is iterated on a space X , say a separable metric space. They term an orbit the set $T^n(x)$, $n = \dots, -1, 0, +1, +2, \dots$, where x is a point of X , define a point x to be periodic if there exists an integer $m > 0$ such that $T^m(x) = x$, and define a point x to be almost periodic if there exists a sequence of integers $n_1 < n_2 < \dots$ such that $\lim_{n \rightarrow \infty} T^{n_i}(x) = x$. This last definition is of course the well known property of (positive) stability in the sense of Poisson and the term almost periodic is somewhat of a misnomer. These topologists are not in the least aware that there is an immense amount of material in dynamics which they should know and they will probably rediscover such interesting things as minimal sets, recurrent motions, minimal centers of attraction, central motions, transitivity, permanent regional transitivity, etc., in the not too distant future. For example, in the last issue of *Mathematical Reviews* (see page 179, review of a paper of Schweigert) Ayres comments that Schweigert has an interesting example for which the periodic points are everywhere dense in a space, but not all the points of the space are periodic. Now this is such a common occurrence in dynamics that we scarcely wonder at it any more.

But I wonder if it is their fault that these things are not better known. For a person who hasn't dealt considerably with these matters it might be a hard task to dig the material out of the literature. One reason is perhaps that in dynamics we deal largely with flows, whereas the topologists deal with the discrete case of a single transformation and its iterates. Though the two are not essentially different, it seems to me that something should be done about this situation. Yet I hesitate about publishing material which can only be considered a rehash of mathematics which is well known (to at least a dozen people).

Fig. 1. Hedlund's letter to Morse: page 1

1940) and *Zentralblatt* (begun about ten years earlier, in 1931) of that period reveals a large number of works concerning periodic and fixedpoint behavior of iterated transformations on manifolds and metric spaces. Of course, there were some more sophisticated precursors: Brouwer's characterization

However, one simple example occurs to me which might make for more awareness concerning the results of dynamics. Let M be the space of symbolic elements of our first paper on symbolic dynamics and let T be the transformation which shifts the index by one, say to the right. Here is a topological transformation which ought to be complex enough to suit the heart of even the most pathological topologist. The periodic trajectories are everywhere dense; the non-periodic recurrent trajectories are everywhere dense; there are transitive trajectories and they form a residual set; the non-periodic, non-recurrent, non-transitive trajectories which are stable in the sense of Poisson are everywhere dense; there are trajectories asymptotic to almost anything; the transformation is permanently regionally transitive. What do you think of giving them this example on which to chew? In view of what we have available in SDI, it should not occupy much space.

Though the preceding example is a good one, it has one defect. The space M is, as we showed, compact, perfect, and totally disconnected. The last property of being totally disconnected should not be essential to the situation and is not characteristic of classical dynamical systems, where the underlying spaces are manifolds. The space M is disconnected because the metric which we chose to topologize the space assumes only discrete values. Would it be possible to topologize the space in some other fashion so that it becomes say a continuum (compact, connected space)? I began thinking of this last night and the answer hasn't occurred to me yet. It should be possible to define a non-trivial metric in the space M of elements such that it goes to zero as larger and larger blocks with center at the preferred symbol become identical and yet such that M is connected. It may be necessary to identify a denumerable set of elements in pairs, but that would not be objectionable. I should think that almost any sort of a space might be obtained with the proper choice of metric, and the whole problem seems interesting.

If all this is of interest to you, I will be glad to learn your reactions.

As ever,

G A H

Fig. 2. Hedlund's letter to Morse: page 2

of fixedpoint-free transformations of the plane (the "Brouwer translation theorem" [Bro12]), Denjoy's work on flows on the torus or, equivalently, diffeomorphisms of the circle [Den32b, Den32a] (see also [vK35]), and the work of Fatou [Fat06, Fat19, Fat20] and Julia [Jul18] on iteration of rational functions.

Hedlund's reference to "the last issue of Mathematical Reviews" may give the impression that his acquaintance with the work of "some topologists" was second-hand. The paper trail suggests otherwise. The review in question, a one-paragraph review of [Sch40] by Ayres: *Math. Rev.* 2, 179b (MR 3198), begins by referring to a paper by Hall and Schweigert [HS38] which is being generalized by Schweigert in the paper under review. Hall was no stranger to Hedlund. A paper by Hall and Kelley [HK41] which appears in Hedlund's bibliography to [Hed44], was published in 1941, based on a presentation to the American Mathematical Society in September 1939. This paper concerns variants of periodicity for an iterated self-homeomorphism of a compact metric space, including uniform and non-uniform versions of almost-periodicity. Hall and Kelley formulate the notion of a minimal set (which they call "irreducibly fixed") and show that this is equivalent to every orbit being dense, and that either such a set is a single periodic orbit, or every orbit in it is almost periodic. In a footnote, they acknowledge that "It has been pointed out to the authors that...[these results]...are precisely analogous to certain results of G. D. Birkhoff for continuous flows..." In an earlier footnote, they acknowledge

This paper was started when the authors were in residence at the University of Virginia, the first named author as a National Research Fellow.

Recall that Hedlund's letter was written in Charlottesville, Virginia, where he had joined the University of Virginia faculty in 1939.

In any case, Morse failed to respond to Hedlund's letter in any substantial way (Figure 3).

Hedlund wrote a new article, "Sturmian minimal sets" [Hed44], submitted to the journal in January, 1944. The minimal sets of the title are addressed in the second half of the paper. Hedlund begins boldly, explicitly introducing the notion of an orbit (note: not "trajectory") and semi-orbit for a discrete dynamical system. He formulates discrete dynamical system versions of the definitions in [Bir27] of α - and ω -limit sets, invariant sets, minimal sets, recurrent orbit (in the sense of being contained in a minimal set), and almost-periodic orbit, and notes the equivalence of the last two notions. Then he comments that the dual terminology "recurrent" and "almost-periodic" for equivalent notions is redundant; he argues that the latter is the better terminology for this notion, and suggests that "recurrent" be saved for "Poisson stable" (our current meaning of "recurrent"). He then repeats the definitions of symbolic trajectory and symbolic element from [MH38], and introduces a modified version of the metric on symbolic elements (replacing m with $m + 1$, so that elements which agree only at the distinguished position are at distance 1), noting that this gives the space of

THE INSTITUTE FOR ADVANCED STUDY
SCHOOL OF MATHEMATICS
PRINCETON, NEW JERSEY

October 6, 1941

Dear Arnold:

We should like to have you come to Princeton to lecture before the Mathematics Club on either the last Wednesday of October or the first Wednesday of November. I think I can obtain your railroad expenses. Louise and I should be very happy to have you stay at our house.

Kakutani, Ambrose, Doob and others are interested in the transitivity problem, and I should like to have you present our joint theorem on topological transitivity for the case of $p > 1$.

The matter you spoke of last spring is of importance and interest. The reason I have not written is because I hoped to have time to think it over first. Unfortunately I have not had time.

Please let me know as soon as possible that you can come, and for which date.

Cordially,

Marston Morse

Professor Gustav A. Hedlund
Mathematics Department
University of Virginia
University, Va.
MM:GB

Marston Morse

Fig. 3. Morse's response

symbolic elements the topology of a Cantor set—apparently abandoning his concerns in the letter about this.

Then, significantly, Hedlund introduces the shift map S and proves that it is a one-to-one, continuous transformation of the space of symbolic elements to itself. He notes that the symbolic trajectories are in one-to-one correspondence with the orbits of S , with periodic trajectories corresponding to periodic orbits, and recalls the existence of the Morse sequence and Sturmian minimal trajectories of [MH40] as examples of almost-periodic non-periodic orbits. He then proceeds to construct Sturmian minimal sets by coding orbits of a rotation of the circle by β radians, where β is an irrational multiple of π , using a partition into arcs of length β and $2\pi - \beta$, respectively ⁽¹²⁾. He associates to each orbit two symbolic elements, corresponding to making the atoms of the partition right- or left-open. He proceeds to prove first that the symbolic sequences which arise this way correspond to almost-periodic orbits of the rotation, and then that this corresponds to being minimal under S . He proves that the minimal set (for S) so obtained is compact, perfect and totally disconnected, and contains a pair of doubly asymptotic orbits. He then defines the notion of an orbit-preserving transformation with respect to a given discrete dynamical system (this is our notion of a self-conjugacy) and proves that in his minimal sets it is not always possible to find an orbit-preserving transformation taking one orbit in the set to an arbitrarily designated second orbit in the set. He attributes the corresponding question for flows to Birkhoff. He defines a notion of almost-periodicity for a transformation (as opposed to a single orbit) and shows that the restriction of S to his minimal sets does not have this property. However, he then defines a notion of *local* almost-periodicity, and shows that his minimal sets do have this property. Finally, he defines a minimal set to be *powerfully minimal* ⁽¹³⁾ if it is minimal under all non-zero iterates of the discrete dynamical system, and proves that his minimal set has this property.

While the focus in this paper is on a specific class of minimal sets (the “Sturmian” ones), Hedlund’s letter suggests that these are now a case study for a more general, abstract study of minimal sets and dynamical properties of the discrete dynamical system defined by S on the space of symbolic elements, and thus open the door to the branch of topological dynamics we now call “symbolic dynamics”.

⁽¹²⁾ Hedlund here works with translation on the real line, but his coding depends only on the mod 1 positions of points.

⁽¹³⁾ Now called *totally minimal*.

REFERENCES

- [AW67] R. Adler and B. Weiss, *Entropy, a complete metric invariant for automorphisms of the torus*, Proc. Nat. Acad. Sci. U.S.A. 57 (1967), 1573–1576.
- [AW70] —, —, *Similarity of automorphisms of the torus*, Mem. Amer. Math. Soc. 98 (1970).
- [Art24] E. Artin, *Ein mechanisches System mit quasiergodischen Bahnen*, Abh. Math. Sem. Hamburg. Univ. 3 (1924), 170–175; reprinted in [Art65, pp. 499–504].
- [Art65] —, *Collected Papers*, Springer, 1965, ed. by S. Lang and J. Tate.
- [Ber67] K. Berg, *On the conjugacy problem for K -systems*, Ph.D. thesis, Univ. Minnesota, 1967.
- [Bes32] A. S. Besicovitch, *Almost Periodic Functions*, Cambridge Univ. Press, 1932; Dover reprint, 1954.
- [Bir12] G. D. Birkhoff, *Quelques théorèmes sur le mouvement des systèmes dynamiques*, Bull. Soc. Math. France 40 (1912), 305–323; reprinted in [Bir50, Vol. 1, pp. 654–672].
- [Bir17] —, *Dynamical systems with two degrees of freedom*, Trans. Amer. Math. Soc. 18 (1917), 199–300; reprinted in [Bir50, Vol. 2, pp. 1–102].
- [Bir20] —, *Surface transformations and their dynamical applications*, Acta Math. 43 (1920), 1–119; reprinted in [Bir50, Vol. 2, pp. 111–229].
- [Bir27] —, *Dynamical Systems*, AMS Colloq. Publ. 9, Amer. Math. Soc., 1927; reissued, revised, 1966.
- [Bir35] —, *Nouvelles recherches sur les systèmes dynamiques*, Mem. Pont. Acad. Sci. Novi Lyncaei 1 (1935), 85–216; reprinted in [Bir50, Vol. 2, pp. 530–661].
- [Bir50] —, *Collected Mathematical Papers*, Amer. Math. Soc., 1950; Dover reprint, 1968.
- [Bro12] L. E. J. Brouwer, *Beweis des ebenen Translationssatz*, Math. Ann. 72 (1912), 37–54; reprinted in [Bro76, Vol. 2, pp. 250–268].
- [Bro76] —, *Collected Works*, North-Holland, 1976, ed. by H. Freudenthal.
- [Den32a] A. Denjoy, *Sur les caractéristiques à la surface du tore*, C. R. Acad. Sci. Paris 194 (1932), 830–833, 2014–2016.
- [Den32b] —, *Sur les courbes définies par les équations différentielles à la surface du tore*, J. Math. Pures Appl. 11 (1932), 5–8.
- [Fat06] P. Fatou, *Sur les solutions uniformes de certaines équations fonctionnelles*, C. R. Acad. Sci. Paris 143 (1906), 546–548.
- [Fat19] —, *Sur les équations fonctionnelles*, Bull. Soc. Math. France 47 (1919), 161–271.
- [Fat20] —, *Sur les équations fonctionnelles*, *ibid.* 48 (1920), 33–94, 208–314.
- [Had98] J. Hadamard, *Les surfaces à courbures opposées et leurs lignes géodésiques*, J. Math. Pures Appl. 4 (1898), 27–73; reprinted in [Had68, Vol. 2, pp. 729–775].
- [Had68] —, *Œuvres de Jacques Hadamard*, Centre National de la Recherche Scientifique, 1968.
- [HK41] D. W. Hall and J. L. Kelley, *Periodic types of transformations*, Duke Math. J. 8 (1941), 625–630.
- [HS38] D. W. Hall and G. E. Schweigert, *Properties of invariant sets under pointwise periodic homeomorphisms*, *ibid.* 4 (1938), 719–724.
- [Hed32a] G. Hedlund, *Geodesics on a two-dimensional Riemannian manifold with periodic coefficients*, Ann. of Math. 33 (1932), 719–739.

- [Hed32b] G. Hedlund, *Poincaré's rotation number and Morse's type number*, Trans. Amer. Math. Soc. 34 (1932), 75–97.
- [Hed35a] —, *A metrically transitive group defined by the modular group*, Amer. J. Math. 57 (1935), 668–678.
- [Hed35b] —, *On the metrical transitivity of the geodesics on closed surfaces of constant negative curvature*, Ann. of Math. 35 (1935), 787–808.
- [Hed36a] —, *Fuchsian groups and transitive horocycles*, Duke Math. J. 2 (1936), 530–542.
- [Hed36b] —, *Two-dimensional manifolds and transitivity*, Ann. of Math. 37 (1936), 534–542.
- [Hed39] —, *The dynamics of geodesic flows*, Bull. Amer. Math. Soc. 451 (1939), 241–260.
- [Hed44] —, *Sturmian minimal sets*, Amer. J. Math. 66 (1944), 605–620.
- [Hil01] D. Hilbert, *Über Flächen konstanter Gausscher Krümmung*, Trans. Amer. Math. Soc. 2 (1901), 87–99.
- [Hir76] M. W. Hirsch, *Differential Topology*, Grad. Texts in Math. 33, Springer, 1976.
- [Hop37] E. Hopf, *Ergodentheorie*, Ergeb. Math. Grenzgeb. 5, Springer, 1937; reissued by Chelsea, 1948.
- [Jor66a] C. Jordan, *Contours tracés sur les surfaces*, J. Math. Pures Appl. 11 (1866), 110–130.
- [Jor66b] —, *La déformation des surfaces*, *ibid.* 11 (1866), 105–109.
- [Jul18] G. Julia, *Mémoire sur l'itération des fonctions rationnelles*, *ibid.* 4 (1918), 47–245.
- [vK35] E. R. van Kampen, *The topological transformations of a simple closed curve into itself*, Amer. J. Math. 57 (1935), 142–152.
- [KU07] S. Katok and I. Ugarcovici, *Symbolic dynamics for the modular surface and beyond*, Bull. Amer. Math. Soc. 44 (2007), 87–132.
- [Kit98] B. P. Kitchens, *Symbolic Dynamics: One-Sided, Two-Sided and Countable State Markov Shifts*, Springer, 1998.
- [Koe31] P. Koebe, *Riemannsche Mannigfaltigkeiten und nichteuklidische Raumforme*, Stizungsber. Preuss. Akad. Wiss. 1927, 345–442; 1928, 414–457; 1929, 304–364, 1930, 505–541; 1931, 506–534.
- [LM95] D. Lind and B. Marcus, *An Introduction to Symbolic Dynamics and Coding*, Cambridge Univ. Press, 1995.
- [Mor21a] M. Morse, *A one-to-one representation of geodesics on a surface of negative curvature*, Amer. J. Math. 43 (1921), 33–51; reprinted in [Mor81, pp. 1–20].
- [Mor21b] —, *Recurrent geodesics on a surface of negative curvature*, Trans. Amer. Math. Soc. 22 (1921), 84–100; reprinted in [Mor81, pp. 21–38].
- [Mor24] —, *A fundamental class of geodesics on any closed surface of genus greater than one*, Trans. Amer. Math. Soc. 26 (1924), 25–60.
- [Mor81] —, *Selected Papers*, Springer, 1981, ed. by R. Bott.
- [MH38] M. Morse and G. Hedlund, *Symbolic dynamics*, Amer. J. Math. 60 (1938), 815–866; reprinted in [Mor81, pp. 443–494].
- [MH40] —, *Symbolic dynamics II. Sturmian trajectories*, *ibid.* 62 (1940), 1–42.
- [Nem49] V. V. Nemyckii [V. V. Nemytskii], *Topological problems of the theory of dynamical systems*, Uspekhi Mat. Nauk 4 (1949), no. 6(34), 91–153 (in Russian); English transl.: Amer. Math. Soc. Transl. no. 103, Amer. Math. Soc., 1954.
- [Nie27] J. Nielsen, *Untersuchungen zur Topologie der geschlossenen zweiseitigen Flächen*, Acta Math. 50 (1927), 189–358; English transl. by J. Stillwell, [Nie86, Vol. 1, pp. 223–341].

- [Nie86] J. Nielsen, *Collected Mathematical Papers*, Birkhäuser, 1986, ed. by V. L. Hansen.
- [Par66] W. Parry, *Symbolic dynamics and transformations of the unit interval*, Trans. Amer. Math. Soc. 122 (1966), 368–378.
- [Poi97] H. Poincaré, *Les méthodes nouvelles de la mécanique céleste*, Gauthier-Villars, 1892, 1893, 1897; Dover reprint, 1957; also republished by Blanchard, Paris, 1987; English transl.: *New Methods of Celestial Mechanics*, NASA, 1967, also AIP Press, 1992 (ed. by D. Goroff).
- [Sch40] G. E. Schweigert, *A note on the limit of orbits*, Bull. Amer. Math. Soc. 46 (1940), 963–969.
- [Sma65] S. Smale, *Diffeomorphisms with many periodic points*, in: *Differential and Combinatorial Topology*, S. Cairns (ed.), Princeton Univ. Press, 1965, 63–80.
- [Thu12] A. Thue, *Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen*, Kristiana (Oslo) Videnskapselskapets Skrifter 1912, no. 1, 1–67; reprinted in [Thu77, pp. 413–477].
- [Thu77] —, *Selected Mathematical Papers of Axel Thue*, Universitetsforlaget, 1977, ed. by T. Nagell.

Department of Mathematics
Wesleyan University
Middletown, CT 06459, U.S.A.
E-mail: ecoven@wesleyan.edu

Department of Mathematics
Tufts University
Medford, MA 02155, U.S.A.
E-mail: zbigniew.nitecki@tufts.edu

Received 9 November 2006

(4823)