INFORMATIQUE THÉORIQUE ET APPLICATIONS

TOM HEAD

The topologies of sofic subshifts have computable Pierce invariants

Informatique théorique et applications, tome 25, nº 3 (1991), p. 247-254.

<http://www.numdam.org/item?id=ITA_1991__25_3_247_0>

Informatique théorique et Applications/Theoretical Informatics and Applications
(vol. 25, n° 3, 1991, p. 247 à 254)

THE TOPOLOGIES OF SOFIC SUBSHIFTS HAVE COMPUTABLE PIERCE INVARIANTS (*)

by Tom HEAD ([1])

Communicated by Jean-Éric PIN

Abstract. – *R. S. Pierce made a study of a class of compact zero-dimensional metric spaces in the early* 1970'*s. The spaces of this class were said to be "of finite type". With each such space he associated a finitary invariant that characterized the structure of the space. We show that the topology of each sofic subshift S is of finite type in the sense of Pierce. Starting from an automaton that recognizes the language of fragments of S we construct a second automaton which recognizes a regular language that has as its adherence a topological space that is homeomorphic with S. On combining this construction with previous work on the adherences of regular languages, we obtain a procedure for constructing the Pierce invariant of the sofic subshift S. This provides an algorithm for deciding homeomorphism of sofic subshifts and suggests that the underlying homeomorphism extension lemmas developed by Pierce may be applicable to more detailed investigations of sofic subshifts.*

Résumé. – *Au début de* 1970 *R. S. Pierce a accompli une étude sur la classe des espaces du compact métrique de dimension zéro. Il définissait les espaces de cette classe comme « de type fini ». A chacun de ces espaces il est associé un invariant fini qui caractérise la structure de l'espace. Nous montrons que la topologie de chaque système sofique S est de type fini comme dans le sens de Pierce. En partant d'un automate qui reconnaît le langage des fragments de S nous pouvons construire un deuxième automate qui reconnaît le langage régulier qui a comme l'adhérence un espace topologique homéomorphe avec S. En réunissant cet construction avec le travail précédent sur les adhérences des langages réguliers nous obtenons une procédure pour construire l'invariant de Pierce d'un système sofique S. Cela nous fournit un algorithme qui décide l'homéomorphisme des systèmes sofiques et suggère que les extensions des lemmes d'homéomorphisme développés par Pierce, sont susceptibles de s'appliquer aux systèmes sofiques.*

1. INTRODUCTION

Let Z be the additive group of integers. Let L: $Z \to Z$ be the function $L(z)=z-1$ and R: $Z \to Z$ be the function $R(z)=z+1$. Let A be a finite non-empty set with the discrete topology and A^Z the set of all functions from Z

(*) Received June 1989, revised April 1990.

([1]) Department of Mathematical Sciences, State University of New York, Binghamton, New York 13901, U.S.A.

into A with the product topology. By a *subshift* S we mean a topologically closed subspace of A^Z for which, for each f in S, the composite functions fL and fR are also in S. With any subshift S we associate a subset $\mathrm{Fr}(S)$ of A^*, *the language of fragments of* S, which consists of the null string and, for every f in S and every pair of integers p & q with $p \leqq q$, the string $f(p)f(p+1)\ldots f(q)$. Each such $\mathrm{Fr}(S)$ is *left prolongeable* in the sense that for each x in $\mathrm{Fr}(S)$ there is an a in A for which ax is also in $\mathrm{Fr}(S)$. Likewise $\mathrm{Fr}(S)$ is *right prolongeable* in that for each x in $\mathrm{Fr}(S)$ there is an a in A for which xa is in $\mathrm{Fr}(S)$. A *sofic subshift* is a subshift S for which the language $\mathrm{Fr}(S)$ is regular [1, 8], *i.e.*, $\mathrm{Fr}(S)$ is recognizable by a finite automaton. The sofic property allows the methodology of formal languages & automata to be applied. We will deal only with sofic subshifts.

L. Boasson & M. Nivat made systematic use of topological concepts in formal language theory in [2] where they defined and studied what is called the adherence of a language. Let A be a finite non-empty set and let L be a subset of A^*. Let P be the set of positive integers. *The adherence of* L is the subset of A^P that consists of all those functions $f\colon P \to A$ for which, for each p in P, the string $f(1)f(2)\ldots f(p)$ occurs as a prefix of a string in L. In [3] it was observed that earlier work of R. S. Pierce [7] allows the complete classification of the topological structure of adherences of regular languages. Pierce associated with an especially simple subclass of the compact zero-dimensional metric spaces, which he defined to be *of finite type*, an invariant structure that completely characterizes these spaces topologically. In [4] it was shown how to construct the Pierce invariant of a regular language from an automaton that recognizes the language. Since the Pierce invariant is a finite structure, to decide whether the adherences of two regular languages are homeomorphic, one may construct and compare these invariants. The next paragraph provides the terminology that is required for a precise definition of the concept of a topological space of finite type.

Let S be a topological space. Let $P(S)$ be the set of all subsets of S. Regards $P(S)$ as an algebra with six operations: φ, S, U, ^{c}, ${}^-$, ${}'$. The first four operations are the usual Boolean operations: φ and S are the nullary operations specifying the minimal & maximal elements for $P(S)$, U is the binary operation of union, and ^{c} is the unary operation of complementation. The final two are unary with ${}^-$ being the operation of topological closure and ${}'$ being the topological derivative. By a subalgebra of $P(S)$ we mean a family of subsets of S that is closed under the six operations of $P(S)$. From the nature of nullary operations every subalgebra must contain both φ & S. Since the intersection of every collection of subalgebras is a subalgebra, $P(S)$ contains a unique minimal subalgebra that necessarily contains $\{\varphi, S\}$. *Let* $\mathrm{Alg}(S)$ *be the minimal subalgebra of* $P(S)$.

DEFINITION: A topological space S is *of finite type in the sense of R. S. Pierce* if the minimal subalgebra Alg (S) is finite.

F. Blanchard & G. Hansel [1] attended the presentation [3] and suggested that there might be a characterization of the topologies of sofic subshifts that is similar to the characterization of the topologies of the adherences of regular languages. The present article has arisen from their suggestion. We indicate the relevance of Pierce's work to the study of sofic systems by showing that each sofic system is homeomorphic with the adherence of the language recognized by an automaton that can be constructed from any automaton that recognizes the language of fragments of the sofic system. This is done in the succeeding sections. If the results presented here suggest to the reader that further study along these lines may be valuable then the homeomorphism extension results of [6] are recommended. That the results of [4] are essentially graph theoretic was shown in [5].

There are two distinct uses of the phrase "of finite type". Pierce's use as defined above and the standard sense used in dynamical systems theory when one speaks of a "subshift of finite type" [1, 8]. *From the present article it follows not only that every subshift of finite type in this latter sense has a topology which is of finite type in the sense of Pierce but that the same holds for every sofic subshift.*

2. THE ISOMETRY H OF A^Z ONTO $(A \times A)^P$

Let A be a finite non-empty set. Let Z be the set of integers. Then A^Z is the set of all functions $f: Z \to A$. We will provide a topology for A^Z by defining a metric d. In preparation for this definition we define, for each f in A^Z and each positive integer p, f_p: $[1-p, p] \to A$ to be the restriction of f to $[1-p, p]$. Thus $f_p(z)=f(z)$ for each z in $[1-p, p]$. For f & g in A^Z we define $d(f, g)=0$ if $f=g$, otherwise $d(f, g)=(1/2)^p$ where p is the least positive integer for which f_p & g_p are distinct. With d as distance function, A^Z is a metric space. The topology of A^Z determined by the metric d coincides with the product topology and therefore, by the theorem of Tychonoff, A^Z is compact.

Let B be a finite non-empty set. Let P be the set of positive integers. Then B^P is the set of all functions $f: P \to B$. We will provide a topology for B^P by defining a metric D. For f & g in B^P we define $D(f, g)=0$ if $f=g$, otherwise $D(f, g)=(1/2)^p$ where p is the least positive integer for which $f(p)$ & $g(p)$ are distinct. With D as distance function, B^P is a metric space. The topology of B^P determined by the metric D coincides with the product topology and therefore B^P is compact.

Let A be a finite non-empty set and let B be the set $A \times A$ of ordered pairs of elements of A. From the paragraphs above we have two metric spaces: A^Z with the metric d, and B^P with the metric D. These spaces are homeomorphic by the following isometry which will be used throughout our work. *Define* $H: A^Z \to B^P$ *by setting, for each* f *in* A^Z *and each* p *in* P, $H[f](p) = (f(1-p), f(p))$.

Thus H maps each doubly infinite sequence of elements of A:

$$\ldots, f(-i), \ldots, f(-2), f(-1), f(0), f(1), f(2), f(3), \ldots, f(i+1), \ldots$$

into the corresponding sequence of ordered pairs of A:

$$(f(0), f(1)), (f(-1), f(2)), (f(-2), f(3)), \ldots, f(-i), f(i+1)), \ldots$$

The definition of the metric d for A^Z was chosen so that it would be apparent at this point that H is an isometry of A^Z onto B^P.

3. H MAPS EACH SOFIC SUBSHIFT ONTO THE ADHERENCE OF A REGULAR LANGUAGE

Let S be a sofic subshift for which the language of fragments, $\mathrm{Fr}(S)$, is recognized by a finite automaton $M = (A, Q, I, T, E)$, where: A is the alphabet of M; Q is the finite set of states of M; I & T are the subsets of Q consisting of the initial and terminal states, respectively; and E is the finite set of edges of M where each edge is an element of the product set $Q \times A \times Q$. It will not be necessary to assume that M is deterministic. We trim M, *i.e.*, delete any states that cannot be reached from I. (When we delete a state we always delete all edges in which the state occurs.) We cotrim M, *i.e.*, delete any states from which T cannot be reached. Since each subword of each word in $\mathrm{Fr}(S)$ is also in $\mathrm{Fr}(S)$, if we make all states of M both initial and terminal the language recognized will be unchanged. Thus we write $M = (A, Q, Q, Q, E)$. From the right prolongability of $\mathrm{Fr}(S)$ we may delete any state in Q which fails to have an outgoing edge. From the left prolongability of $\mathrm{Fr}(S)$ we may delete any state in Q which fails to have an incoming edge. With the appropriate deletions made we may assume that $\mathrm{Fr}(S)$ is recognized by a trim cotrim finite automaton $M = (A, Q, Q, Q, E)$ for which every state has both an entrance and an exit.

We will construct from $M = (A, Q, Q, Q, E)$ a second finite automaton $M' = (B, Q', D, Q', E')$ for which we can assert that, for the isometry $H: A^Z \to B^P$ defined in Sec. 1, H maps the sofic system S onto the adherence of the language recognized by M'. The alphabet of M' will be the set $B=A \times A$ of ordered pairs of elements of A. The set Q' of states of M' will be a subset, constructed below, of the set $Q \times Q$ of ordered pairs of states of M. The set of initial states of M' will be the diagonal D of $Q \times Q$, *i.e.*, $D=\{(q, q): q \text{ in} Q\}$. All states of M' will be terminal. The set E' of edges will be ordered triples of ordered pairs of the form $[(q(1), q(2)), (a(1), a(2)), (q(3), q(4))]$. We construct Q' and E' together as we carry to completion the following construction that we initiate by taking Q' to be D and E' to be empty. Thus M' is initially (B, D, D, D, φ). We now begin an iterative procedure:

For each state (q, q') of M' (that has not already been treated) and each ordered pair (a, a') in $B=A \times A$ we compute the set of pairs $t=\{(p, p')$ in $Q \times Q$: (p, a, q) & (q', a', p') are edges of $M\}$. If t is not empty we modify M' by adding to Q' each of the pairs (p, p') that is not already in Q' and by adding to E' each of the triples $[(q, q'), (a, a'), (p, p')]$ that is not already in E'.

The procedure described immediately above is carried out until there is no state of M' that has not been so treated. The process must terminate because the number of possible states is finite. As this procedure is repeated, the set Q' of states of M' and the set E' of edges of M' grow until stability is reached at the time when no states remain to be treated. We take this final form of M' to be the automaton we sought and denote it: $M'=(B, Q', D, Q', E')$.

THEOREM: *Let A be a finite non-empty set and let $B=A \times A$. Let $H: A^Z \to B^P$ be the isometry defined in Sec. 2. Let S be any sofic subshift in A^Z. Construct from* $\mathrm{Fr}(S)$ *the finite automaton $M=(A, Q, Q, Q, E)$ and then from M construct $M'=(B, Q', D, Q', E')$ as described above. Then H maps S isometricly onto the adherence of the language recognized by M'.*

COROLLARY 1: *Each sofic subshift is a compact zero-dimensional metric space that is of finite type in the sense of R. S. Pierce.*

COROLLARY 2: *Homeomorphism of sofic subshifts is decidable.*

Corollary 1 follows from the theorem since adherences of regular languages are of finite type [4]. Corollary 2 follows from the decidability of homeomorphism of such adherences. The proof of the theorem will be given in the final section. Here we wish to emphasize the algorithmic yield of the theorem. *If a sofic subshift S is given in a constructive sense, so that an automaton recognizing its fragment language* $\mathrm{Fr}(S)$ *is either given or can be constructed, then the automata M & M' can be constructed. From M' the Pierce* *invariant of S can be constructed as in* [4]. *If two sofic subshifts are given then their Pierce invariants can be constructed and compared to decide whether the subshifts are homeomorphic.*

4. THE REQUIRED PROOFS

For each string x in $\mathrm{Fr}(S)$ the automaton $M=(A, Q, Q, Q, E)$ that was constructed in Sec. 3 possesses, of course, a path of finite length that is labeled by x. Of prime interest here is the fact that for each function f in S, M also possesses a doubly infinite path that is labeled by the sequence consisting of the values assumed by the function f. It is this additional representation power of M that allows us to reduce the determination of the topological structure of sofic subshifts to that of adherences of regular languages. We clarify this point by presenting a lemma.

LEMMA 1: *For each f in S there is at least one doubly infinite path in $M=(A, Q, Q, Q, E)$ of the following form:*

$$\ldots [q(-j-1), f(-j), q(-j)][q(-j), f(-j+1), q(-j+1)] \ldots$$
$$[q(-1), f(0), q(0)][q(0), f(1), q(1)] \ldots$$
$$[q(j-1), f(j), q(j)][q(j), f(j+1), q(j+1)] \ldots$$

Proof: For each f in $\mathrm{Fr}(S)$ & each $N>0$, $\mathrm{Fr}(S)$ contains $f(1-N)f(2\ \ N)\ldots f(N-1)f(N)$. As N assumes successive positive values this provides an infinite sequence of string in $\mathrm{Fr}(S)$, one of each even length. Since each of these strings is recognized by $M=(A, Q, Q, Q, E)$, there is, for each $N>0$, an acceptance path:

$$[q(N, -N), f(1-N), q(N, 1-N)][q(N, 1-N), f(2-N), q(N, 2-N)] \ldots$$
$$[q(N, N-2), f(N-1), q(N, N-1)][q(N, N-1), f(N), q(N, N)],$$

where the various $q(N, K)$ are states of M. As N assumes successive positive values we have an infinite sequence of paths in M, one of each even length. In order to produce a doubly infinite path having the properties we desire, we must choose an appropriate subsequence of this sequence. This will be done by repeatedly using the fact that *there are only finitely many paths of any one given length in M*:

For every $N>0$, the path of length $2N$ contains a "subpath" $[q(N, -1), f(0), q(N, 0)][q(N, 0), f(1), q(N, 1)]$ of length 2. *We will call this specific subpath the middle subpath of length* 2 *in the path of length* $2N$. Thus we have an infinite sequence, as N varies, of middle subpaths of length 2. Since there are only finitely many distinct paths of length 2 in M, there exists at least one path $[q, f(0), q'][q', f(1), q'']$ of length 2 that occurs infinitely often as the middle subpath of length 2. Thus in the original sequence of paths, one of each even length, there is at least one infinite subsequence for which all the middle subpaths of length 2, $[q(N, 1), f(0), q(N, 0)][q(N, 0), f(1), q(N, 1)]$ coincide. Delete all the paths in the original sequence of paths except those that have this common middle subpath of length 2. $\{$Notice that, after passing to the chosen subsequence, $q(N, -1)=q(K, -1)$ & $q(N, 0)=q(K, 0)$ & $q(N, 1)=q(K, 1)$ *hold for all remaining positive integers* N & K. Thus we may lighten the notation by writing $q(-1)=q(N, -1)$ & $q(0)=q(N, 0)$ & $q(1)=q(N, 1)$.$\}$ The step taken in this paragraph can be repeated indefinitely:

In the infinite sequence of even length paths *that remain* there must be an infinite subsequence of paths that share *the same middle subpath of length* 4. Delete all the paths of the sequence except those that have this common middle subpath of length 4. $\{$Lighten the notation by using $q(-2)$ & $q(2)$.$\}$

Continue from one subsequence to the next sub-subsequence. In the process we generate "from the center" successive pairs of edges labeled by the desired values of the function f. $\square$

From Lemma 1 and the method by which M' was constructed from M, it follows that $H(S)$ is contained in the adherence of the language recognized by M':

Let f be any function in S. Then M contains at least one doubly infinite path as in the statement of Lemma 1. From the construction of M' it follows that M' contains the following unending path:

$$[(q(0), q(0)), (f(0), f(1)), (q(-1), q(1))]$$
$$[(q(-1), q(1)), (f(-1), f(2)), (q(-2), q(2))]\ldots$$
$$[(q(-j), q(j)), (f(-j), f(j+1)), (q(-j-1), q(j+1))]\ldots$$

This assures that $H(f)=(f(0), f(1))(f(-1), f(2))\ldots(f(-j), f(j+1))\ldots$ is in the adherence of the language recognized by M'.

The following lemma provides the final fact needed:

LEMMA 2: *For each sequence t in the adherence of the language recognized by M' there is an f in S for which $H(f)=t$.*

Proof: Let $t=(a(1), a(2))(a(3), a(4))\ldots(a(N), a(N+1))\ldots$ be a sequence that lies in the adherence of the language recognized by M'. Since every state of M' is terminal it follows that every finite prefix, $(a(1), a(2))\ldots(a(N), a(N+1))$, of t is recognized by M'. Thus for each $N>0$ there is an acceptance path: $[(q(N, 0), q(N, 0)), (a(1), a(2)), (q(N, 1),$

$q(N, 2))] [(q(N, 1), q(N, 2)), (a(3), a(4)), (q(N, 3), q(N, 4))] \ldots [(q(N, N-2), q(N, N-1)), (a(N), a(N+1)), (q(N, N), q(N, N+1))]$. As N assumes successive positive values we obtain an infinite sequence of such paths. We now proceed as we did in the proof of Lemma 1. We use the fact that *there are only finitely many paths of any one given length in M'*. Applying this fact step by step as we did in Lemma 1 we see that there is an unending path in M' of the form:

$$[(q(0), q(0)), (a(1), a(2)), (q(1), q(2))]$$
$$[(q(1), q(2)), (a(3), a(4)), (q(3), q(4))] \ldots$$
$$[(q(N-2), q(N-1)), (a(N), a(N+1)), (q(N), q(N+1))] \ldots$$

By the construction of M' it follows that there is in M the doubly infinite path:

$$\ldots [q(N), a(N), q(N-2)] \ldots [q(3), a(3), q(1)]$$
$$[q(1), a(1), q(0)] [q(0), a(2), q(2)] [q(2), a(4), q(4)] \ldots$$
$$[q(N-1), a(N+1), q(N+1)] \ldots$$

The presence of this doubly infinite path in M assures us that the function $f: Z \to A$, defined for $z > 0$ by $f(z) = a(2z)$ and for $z \leqq 0$ by $f(z) = a(1-2z)$, lies in the subshifts S and that $H(f) = t$. $\square$

REFERENCES

1. F. Blanchard and G. Hansel, Languages and Subshifts, *in* M. Nivat and D. Perrin Eds., *Automata on Infinite Words*, pp. 138-146, Lectures Notes in Comput. Sci., *192*, Springer-Verlag, 1985.
2. L. Boasson and M. Nivat, Adherences of Languages, *J. Comput. Syst. Sci.*, 1980, *20*, pp. 285-309.
3. T. Head, The Adherences of Languages as Topological Spaces, *in* M. Nivat and D. Perrin Eds., *Automata on Infinite Words*, pp. 147-163, Lecture Notes in Comput. Sci., *192*, Springer-Verlag, 1985.
4. T. Head, The Topological Structure of Adherences of Regular Languages, *RAIRO, Inform. Théor. Appl.*, 1986, *20*, pp. 31-41.
5. T. Head, The Topological Structure of the Space of Unending Paths of a Graph, *Congr. Numer.*, 1987, *60*, pp. 131-140.
6. R. S. Pierce, Existence and Uniqueness Theorems for Extensions of Zero-Dimensional Metric Spaces, *Trans. Amer. Math. Soc.*, 1970, *148*, pp. 1 -21.
7. R. S. Pierce, Compact Zero-Dimensional Metric Spaces of Finite Type, *Mem. Amer. Math. Soc.*, No. 130, Providence, Rhode Island, 1972.
8. B. Weiss, Subshifts of Finite Type and Sofic Systems, *Monatsh. Math.*, 1973, *77*, pp. 462-474.