SOFIC CONSTANT-TO-ONE EXTENSIONS OF SUBSHIFTS OF FINITE TYPE

F. BLANCHARD AND G. HANSEL

(Communicated by Dennis Burke)

ABSTRACT. Provided a constant-to-one extension of a subshift of finite type is sofic and transitive, then it is also of finite type.

1. INTRODUCTION

In the setting of metric ergodic theory, Rudolph [7, 8] proved that any weakly mixing, constant-to-one extension of a Bernoulli system is also Bernoulli. Constant-to-one extensions have also been given some attention in symbolic dynamics: in [6], Nasu investigated constant-to-one factor maps between subshifts of finite type and proved that they are necessarily right- and left-closing. In [1], one of the authors has shown that transitive constant-to-one extensions of coded systems (defined in [2]) are coded or equivalently, according to Krieger's characterization, are the limit of an increasing sequence of transitive subshifts of finite type. The corresponding statement for sofic systems was proved to be false, also in [1].

In this paper, we address two questions.

Question 1. Let S be a transitive, constant-to-one extension of the subshift of finite type T . Is S also of finite type?

We do not give the general answer to this question. We prove that if, moreover, S is assumed to be sofic, then it is indeed of finite type (§3). The bulk of the proof consists of showing that in this situation, the factor map f must be closing. This allows us to make use of an (unstated) result of Kitchens, the proof of which is to be found in [3]: any closing transitive extension of a subshift of finite type is of finite type.

It would be nice if any transitive, constant-to-one extension of a sofic system were sofic, because if this were true Question 1 would be completely solved. Unfortunately, this is false, as remarked above.

Considering Nasu's result for constant-to-one factor maps between transitive

Received by the editors October 6, 1989 and, in revised form, April 17, 1990.

1980 *Mathematics Subject Classification* (1985 Revision). Primary 54H20; Secondary 28D05.

© 1991 American Mathematical Society
0002-9939/91 \$1.00 + \$.25 per page

subshifts of finite type [6], the following question arises naturally:

Question 2. Let S and T be two transitive sofic systems and $f: S \rightarrow T$ be a constant-to-one factor map. Need f be right- and left-closing?

The answer is no, as shown by a counterexample (§4).

2. NOTATIONS AND DEFINITIONS

Let A be a finite alphabet. A^* is the set of words on A , i.e., the set of finite sequences of elements of A , and $A^{\mathbb{Z}}$ the set of bi-infinite sequences $(a_n)_{n \in \mathbb{Z}}$ of elements of A . The length of a word $u \in A^*$ is denoted by $|u|$.

Let $s = (a_n)_{n \in \mathbb{Z}} \in A^{\mathbb{Z}}$. For $m \leq n$, set

$$s[m, n] = a_m a_{m+1} \cdots a_n.$$

The *past* (the *future*) of the word s is the left-infinite sequence $s^- = s[] - \infty, 0[] = (a_n)_{n < 0}$ (the right-infinite sequence $s^+ = s[0, +\infty[] = (a_n)_{n \geq 0}$); one writes $s = (s^-, s^+)$. A *periodic point* $s \in S$ is a point for which there exists an integer $p > 0$ such that $a_{n-p} = a_n$ for any $n \in \mathbb{Z}$; a *periodic past* is one which fulfills the same condition for $n < 0$; in the two cases p is called a *period* of s or s^- .

Let L be a subset of A^* ; a word $v \in L$ is *periodic in L* if, for any $n \in \mathbb{N}$, the word v^n still belongs to L . Then the infinite periodic sequence $s = (a_n)_{n \in \mathbb{Z}}$, obtained by concatenating v with itself infinitely many times, such that $a_0 a_1 \cdots a_{|v|-1} = v$ is denoted ${}^\omega v$; its past s^- is denoted ${}^\omega v$.

A *subshift* is a closed shift-invariant subset of the compact set $A^{\mathbb{Z}}$.

Given a subshift $S \subset A^{\mathbb{Z}}$, the set $L(S)$ of all words which are factors of elements of S , i.e.,

$$L(S) = \{s[m, n] \mid s \in S, m, n \in \mathbb{Z}, m \leq n\},$$

is called the *language* of S . The subshift S is said to be a *subshift of finite type* if it consists of the set of all bi-infinite sequences $s = (a_n)$ all finite factors $s[m, n]$, $m \leq n$ of which do not belong to some given finite set H of excluded words. A subshift S is said to be *transitive* if for any $u, v \in L(S)$, there exists $w \in A^*$ such that $uwv \in L(S)$.

Let $S \subset A^{\mathbb{Z}}$ and $T \subset B^{\mathbb{Z}}$ be two subshifts. A *factor map* from S to T is a continuous, onto, shift-commuting mapping $f: S \rightarrow T$. In this situation T is a *factor* of S and S is an *extension* of T . A *one-block* factor map $f: S \rightarrow T$ is one for which there exists a map $\phi: A \rightarrow B$, onto and shift-commuting, such that $f((a_n)_{n \in \mathbb{Z}}) = (\phi(a_n))_{n \in \mathbb{Z}}$ for any $(a_n) \in S$. In this case we shall write, improperly, $f(a_1 \cdots a_n)$ instead of $\phi(a_1) \cdots \phi(a_n)$, for the sake of simplicity. Two subshifts S and T are *conjugate* if there exists a one-to-one factor map from S onto T . *Sofic systems* are all subshifts which are images of subshifts of finite type under factor maps. Any subshift conjugate to a subshift of finite type (sofic system) is also of finite type (sofic). Moreover, any factor map $f: S \rightarrow T$ may be rendered one-block by substituting conjugate subshifts to S and T , if needed. One gets an important characterization of sofic systems by the use

of automata (in a restricted sense): an *automaton* on alphabet A is a couple $\mathcal{A} = (E, G)$ with E a finite set of *states* and G a subset of the cartesian product $Q \times A \times Q$. The elements of G are the *edges* of the automaton. Let us also assume that the automaton is deterministic; i.e., for any $(e, a) \in E \times A$, there exists at most one state e' such that $(e, a, e') \in G$. A *path* in $\mathcal{A}$ is a sequence (e_i, a_i, e_{i+1}) , $i = 0, 1, \dots, n$, of consecutive edges. Its *label* is the word $w = a_1 a_2 \dots a_n$. A word $w \in A^*$ is *accepted* by the automaton $\mathcal{A}$ if there exists a path with label w . The language *recognized by the automaton* is the set of all accepted words. The following holds (see for instance [4]):

A subshift S is sofic iff the language $L(S)$ is recognized by an automaton.

Notice that, as E is finite, it is equivalent to say that, for any $s = (a_n)_{n \in \mathbb{Z}} \in S$, there exists a sequence of states $(e_n)_{n \in \mathbb{Z}}$ such that $(e_n, a_n, e_{n+1}) \in G$ for any $n \in \mathbb{Z}$.

3. SOFIC CONSTANT-TO-ONE EXTENSIONS

Let S and T be two subshifts (on alphabets A and B). A factor map $f: S \rightarrow T$ is *bounded-to-one*, actually *at most k -to-one* in the present case (*constant-to-one*, here *k -to-one*) if there exists a number k such that for any $t \in T$, $\text{card}(f^{-1}(t)) < k$ ($\text{card}(f^{-1}(t)) = k$). The factor map $f: S \rightarrow T$ is defined to be *right-closing* if for $s, s' \in S$, $s^- = s'^-$ and $f(s) = f(s')$ imply $s = s'$.

The following result, due to Kitchens, may be found in [3, second proof of Proposition 4.12].

Proposition 1. *Let S be a transitive subshift, T a subshift of finite type and $f: S \rightarrow T$ a right-closing factor map. Then S is also a subshift of finite type.*

Remark. It follows from Proposition 1 that a right-closing transitive extension of a sofic system is itself sofic. Indeed let S' be a right-closing transitive extension of the sofic system S with the factor map $\psi: S' \rightarrow S$. Let $\phi: T \rightarrow S$ be a right-closing factor map, where T is a transitive subshift of finite type (it is well known that such a factor map exists). Let $f: T \times S' \rightarrow T$ and $g: T \times S' \rightarrow S'$ be the canonical projection maps. Let $(t, s') \in T \times S'$ be a point such that $\phi(t) = \psi(s')$ has a dense orbit in S . Let Σ be the orbit closure of (t, s') in $T \times S'$. Then the system Σ is transitive, and the restriction of f to Σ is a right-closing factor map. Hence, by Proposition 1, Σ is a subshift of finite type. Since S' is the image of Σ under g , it is a sofic system.

The next result is a characterization of non-right-closing factor maps when S is sofic.

Proposition 2. *Let $f: S \rightarrow T$ be a factor map, and assume S to be sofic. Then the following conditions are equivalent:*

- (1) f is not right-closing;

- (2) *there exist two distinct points s and s' in S , such that $f(s) = f(s')$, having the same periodic past.*

Proof. (2) $\Rightarrow$ (1) is obvious. Let us show (1) $\Rightarrow$ (2). The hypothesis means that we can find two distinct points $s = (a_n)$ and $s' = (a'_n)$ in S such that $f(s) = f(s')$ and $s^- = s'^-$. Let $\mathcal{A} = (E, G)$ be an automaton recognizing the language $L(S)$. There exist two sequences of states $(e_n)_{n \in \mathbb{Z}}$ and $(e'_n)_{n \in \mathbb{Z}}$ in $\mathcal{A}$ such that for any $n \in \mathbb{Z}$, (e_n, a_n, e_{n+1}) and (e'_n, a'_n, e'_{n+1}) belong to G . As the set E is finite, there exist two integers $m < p < 0$ such that

$$e_m = e_p \quad \text{and} \quad e'_m = e'_p.$$

Set

$$u = a_m a_{m+1} \cdots a_{p-1} = a'_m a'_{m+1} \cdots a'_{p-1}.$$

The word u is periodic in $L(S)$ and the sequences

$$s_1 = ({}^\omega u; (a_p a_{p+1} \cdots)) \quad \text{and} \quad s_2 = ({}^\omega u; (a'_p a'_{p+1} \cdots))$$

answer the question.

Let $f: S \rightarrow T$ be a one-block factor map, and let (u, v, w) be a triple of words such that $uvw \in L(T)$. The word v' is a *lift of v in context (u, w)* if there exist words u' and w' such that

$$|u'| = |u|, \quad |w'| = |w|, \quad u'v'w' \in S \quad \text{and} \quad f(u'v'w') = uvw.$$

When $f(v') = v$, v' is merely called a *lift of v* .

Proposition 3. *Let $f: S \rightarrow T$ be an at most k -to-one factor map.*

- (1) *For any $n \in \mathbb{N}$, there exists $m = m(n)$ such that, for any triple of words (u, v, w) such that $|u| = |w| = m$, $|v| = n$, there exist at most k distinct lifts of v in context (u, w) .*
- (2) *Assuming S to be sofic, $m(n)$ may be chosen constant.*

Proof. (1) relies on a simple compactness argument. Suppose (2) is false, and let p be the number of states of an automaton $\mathcal{A}$ recognizing $L(S)$. According to the hypothesis, there exists a triple of words (u, v, w) such that v has more than k distinct lifts in context (u, w) and

$$(1) \quad |u| = |w| = p^{k+1} + 1.$$

Let (u_i, v_i, w_i) , $i = 1, \dots, k+1$, be a sequence of $k+1$ triples which are lifts of (u, v, w) , with the v_i all distinct and such that $u_i v_i w_i \in L(S)$. Let us show that one can find, “simultaneously for all i ,” a factor u'_i of u_i and a factor w'_i of w_i which may be repeated any number of times without diminishing the number of lifts of v in the corresponding context, which contradicts (1). More precisely, for any $i = 1, \dots, k+1$, let

$$(e_0^i, e_1^i, \dots, e_{|u|}^i, e_{|u|+1}^i, \dots, e_{|u|+|v|}^i, e_{|u|+|v|+1}^i, \dots, e_{|u|+|v|+|w|}^i)$$

be the sequence of states in a path with label $u_i v_i w_i$ in the automaton $\mathcal{A}$. According to (1), there exist integers r, s, r', s' such that $0 \leq r < s \leq |u|$, $|u| + |v| \leq r' < s' \leq |u| + |v| + |w|$, and

$$e_r^i = e_s^i, \quad e_{r'}^i = e_{s'}^i, \quad i = 1, \dots, k+1.$$

As a consequence, the words u_i and w_i have factorizations $u_i = x_i u'_i y_i$ and $w_i = t_i w'_i z_i$ such that

$$\begin{aligned} |x_1| &= \dots = |x_{k+1}|, & |u'_1| &= \dots = |u'_{k+1}|, \\ |t_1| &= \dots = |t_{k+1}|, & |w'_1| &= \dots = |w'_{k+1}|, \end{aligned}$$

and that, for any $n \in \mathbb{N}$,

$$x_i (u'_i)^n y_i v_i t_i (w'_i)^n z_i \in L(S).$$

Hence there exists a family of words $(\alpha_i v_i \beta_i)$, $i = 1, \dots, k+1$, with the same image, some words in α_i and β_i having arbitrary length, which contradicts (1).

Let $f: S \rightarrow T$ be an at most k -to-1 factor map. A triple of words (u, v, w) such that $uvw \in L(T)$ is said to be *biclosing* if, for any $x, y \in L(T)$ such that $xuvw y \in L(T)$, the word v has exactly k lifts in context (xu, vw) . Obviously, if (u, v, w) is biclosing, any triple (xu, v, wz) such that $xuvwz \in L(T)$ is also biclosing. The next proposition is proved in [1].

Proposition 4. *Let $f: S \rightarrow T$ be a constant-to-one factor map. Then $L(T)$ contains biclosing words.*

Proposition 5. *Let $f: S \rightarrow T$ be a k -to-one one-block factor map. Let v be periodic in $L(T)$. Then there exists $p \in \mathbb{N}$ such that for any $n > 0$, v^p has exactly k distinct periodic lifts in context (v^n, v^n) .*

Proof. Let $s_1, \dots, s_k$ be the lifts of ${}^\omega v^\omega$. They are all periodic with the same period $m = p \cdot |v|$. As the s_i are distinct, they must differ on any interval with length m .

Remark. Assume S is sofic, and let m be the number the existence of which is stated in Proposition 3. Let $v \in L(T)$ be a periodic word such that $|v| > m$. Let p be the integer the existence of which is stated in Proposition 5. Then the word $w = v^p$ has exactly k lifts $v_1, \dots, v_k$ in context (v, v) , and those k lifts are periodic.

Theorem 1. *Let $f: S \rightarrow T$ be a k -to-one factor map with S sofic and transitive and T of finite type. Then f is right-closing.*

Proof. Assume f is not right-closing: then there must exist $s = (\alpha; \beta_1)$ and $s' = (\alpha; \beta_2)$ in S such that α is periodic, and β_1 and β_2 are distinct with the same image under f . As T is of finite type, there exists an integer q such that if c, d, e are three words with cd and de in $L(T)$, and if $|d| \geq q$, then $cde \in L(T)$. Let m be the integer the existence of which is stated in Proposition 3. Let v be such that $f(\alpha) = {}^\omega v$, $|v| > \max(m, q)$ and having k

distinct periodic lifts $v_1, v_2, \dots, v_k$. Hence the lifts of v in context (v, v) are exactly $v_1, \dots, v_k$. For $t = f(s) = f(s')$, let $s_1 = s, s_2 = s', s_3 \cdots s_k$ be the preimages of t . One may assume the s_i^+ are all distinct and the s_i^- are all periodic (if it is not so, it suffices to shift t to the right by a multiple of $|v|$).

As $s_1^- = s_2^- = \alpha$, there are at most $k - 1$ distinct lifts of $f(\alpha)$ among the s_i^- . One may assume hence that

$$(2) \quad s_1^- = s_2^- = {}^\omega v_2$$

and

$$(3) \quad s_i^- = {}^\omega \phi_i, \quad \phi_i \subset \{v_2, v_3, \dots, v_k\}, \quad i = 3, 4, \dots, k.$$

Let p be an integer such that the words $s_i[0, p]$, $i = 1, \dots, k$ are all distinct. The word $d = t[-3|v|, p + m]$ has factorization $d = v^3 y z$ with $|z| = m$. As $|v|$ and $|z|$ are greater than m , the word $v^2 y$ admits at most k distinct lifts in context (v, z) . As a consequence of (2) and (3), those lifts are

$$(4) \quad v_2^2 s_1[0, p], \quad v_2^2 s_2[0, p], \quad \text{and} \quad \phi_i^2 s_i[0, p] \quad \text{for } i = 3, \dots, k.$$

The word v_1^2 is a left factor of none of them.

As S is transitive, there exists a triple (u', x', w') such that $u'x'w'v_1^3 \in L(S)$ and $(u, x, w) = (f(u'), f(x'), f(w'))$ is a biclosing triple, with $|u| > m$. The words $uxwv^3$ and $v^3 y z$ belong to $L(T)$. Since $|v^3| > q$, this is also true for the word $uxwv^3 y z$. The triple $(u, x, wv^3 y z)$ is biclosing because it is an expansion of (u, x, w) . Hence the word $xwv^3 y$ has exactly k lifts in context (u, z) . They have the form $x_i w_i v_i' v_i'' v_i''' y_i$, $i = 1, \dots, k$, with the x_i exactly spanning the k distinct lifts of x in context (u, w) . But for any $i = 1, \dots, k$ the word $v_i'' v_i''' y_i$ is a lift of $v^2 y$ in context (v, z) ; hence, from (4),

$$v_i'' = v_i''' \neq v_1.$$

This implies that the word xwv^2 has $k + 1$ lifts in context (u, v) : $x'w'v_1^2$, and the k words $x_i w_i v_i' v_i''$, $i = 1, \dots, k$. On account of the lengths of u and v , this contradicts Proposition 3.

Corollary. *Let S be a subshift, T a subshift of finite type, and $f: S \rightarrow T$ a constant-to-one factor map. If S is sofic and transitive, then it is also of finite type.*

Proof. It is a direct consequence of Theorem 1 and Proposition 1.

4. A COUNTEREXAMPLE

Here are two transitive sofic systems S and T , with S a nonclosing constant-to-one extension of T . In fact, this is a simpler version of [1, Example 2].

The system T is Weiss's even system: all finite strings of b 's between two a 's must have even length. The extension S is defined on the alphabet $\{a_1, a_2, b_1, b_2\}$ and $f: S \rightarrow T$ just drops subscripts. In order to define S , forbid words

as follows:

- (a) $a_i a_j$, $i \neq j$; i.e., two consecutive a 's have the same label;
- (b) $a_i b_2$, $b_1 a_i$, $b_i b_i$, $i = 1, 2$; i.e., two consecutive b 's have different labels, the first b in a string must be b_1 , the last one b_2 .
- (c) $a_i (b_1 b_2)^{2^p} a_i$, $i = 1, 2$; i.e., a 's must change label when one goes to the next string.

It is quite easy to check that S is sofic and transitive and f is everywhere 2-to-1. But f is not closing, on account of the set of rules (c): for $t \in T$, with $t_0 = a$ and $t_i = b$ for $i < 0$, the subscript of a at time 0 is not determined, whereas there is only one way of labelling all b 's to the left.

ACKNOWLEDGMENTS

The referee made several valuable suggestions. The most important was Question 2. He or she also pointed out that the soficity of transitive closing extensions of sofic systems is a corollary of Proposition 1.

REFERENCES

1. F. Blanchard, *Extensions à fibre constante*, Ergodic Theory Dynamical Systems (to appear).
2. F. Blanchard and G. Hansel, *Systèmes codés*, Theoret. Comput. Sci. **44** (1986), 17–49.
3. M. Boyle, B. Marcus, and P. Trow, *Resolving maps and a dimension group for subshifts of finite type*, Mem. Amer. Math. Soc., no. 377, Amer. Math. Soc., Providence, RI, 1987.
4. R. Fischer, *Sofic systems and graphs*, Monatsh. Math. **80** (1975), 179–186.
5. W. Krieger, personal communication.
6. M. Nasu, *Constant-to-one and onto global maps of homomorphisms between strongly connected graphs*, Ergodic Theory Dynamical Systems **3** (1983), 387–414.
7. D. Rudolph, *If a two-point extension of a Bernoulli shift has an ergodic square, then it is Bernoulli*, Israel J. Math. **30** (1978), 159–180.
8. —, *If a finite extension of a Bernoulli shift has no finite rotation factors, it is Bernoulli*, Israel J. Math. **30** (1978), 193–206.

UNIVERSITÉ DE PROVENCE, UFR MATHÉMATIQUES, CASE X, 3 PLACE VICTOR HUGO, 13331 MARSEILLE CÉDEX 3, FRANCE

DÉPARTEMENT DE MATHÉMATIQUES, FACULTÉ DES SCIENCES DE ROUEN, B.P. 118, 76134 MONT-SAINT-AIGNAN CÉDEX, FRANCE