FACTORING FACTOR MAPS

Mike Boyle

Abstract. A noninjective bounded-to-one factor map from an irreducible shift of finite type onto a sofic system can be factored as a composition of other such maps in only finitely many ways (up to isomorphism). This generalizes to factor maps from systems with canonical coordinates to finitely presented dynamical systems.

Section 0. Introduction

The analysis of bounded-to-one and especially of resolving factor maps has played a key role in the study of various coding relations among shifts of finite type and Markov chains [2, 3, 7, 14, 20 ...]. The classification of sofic systems up to topological conjugacy reduces to the classification of such maps, because sofic systems come with canonical resolving factor maps from shifts of finite type [15, 5]. The study of these maps is also at the heart of applications of symbolic dynamical ideas to the problem of efficiently coding data into and out of magnetic media [1, 12, 17].

The problem of determining when there exists a resolving map between two given irreducible shifts of finite type (SFT's) was solved by Ashley [3]. This problem for bounded-to-one maps is still open, but the necessary conditions of [14] are conjectured [4, p.34] to be sufficient. The only known general constructions of such maps are as resolving maps and their compositions (e.g. [2, 3, 7]).

Let $\mathcal{B}$ be the class of bounded-to-one factor maps from irreducible shifts of finite type onto sofic systems. With the operation of composition (when defined), $\mathcal{B}$ is a category. One natural goal in this setting is to understand the factorizations of maps in $\mathcal{B}$. Adler and Marcus [2] asked if bounded-to-one factor maps between irreducible SFT's are always compositions of resolving maps, a question answered in the negative by Kitchens [14]. Certain sofic systems T (the AFT systems of Marcus [17]) are the images of factor maps which are "minimal covers" through which every factor map in $\mathcal{B}$ onto T must factor, and all irreducible sofic systems have "minimal resolving covers" [5]. Recently Trow [21] proved that a map in $\mathcal{B}$ can have only finitely many factorizations (up to isomorphism) as a left resolving map followed by a right resolving map, and such a factorization is unique (up to isomorphism) if the map has degree 1. His proofs rest on nontrivial constructions of accessory factor maps, generalizing constructions of Nasu [19] and built up via magic words in an elaboration of ideas in [14]. It was this work of Trow which stimulated the current paper.

1991 *Mathematics Subject Classification.* 58F03, 54H20.

This work was supported by NSF Grant DMS 9401538 and the University of Warwick Maths Institute.

In this paper we prove that in $\mathcal{B}$ every element γ has a finite factorization into irreducibles, and there are (up to isomorphism) only finitely many such factorizations for γ, and they can all be computed. It turns out that this can be done quite simply and concretely by studying relations among equivalence relations of these factor maps. The key point is to relate the "no diamonds" property of maps in $\mathcal{B}$ (i.e., these maps do not collapse distinct biasymptotic points) to the equivalence relations. The approach then gives similar results for factor maps from systems with canonical coordinates to finitely presented systems. We also recover in this setting the results of Trow [21] described above.

As one application of the tools developed here, some conditions (2.10-12) are given which rule out the possibility that a system is a quotient of an SFT. As another we prove a proposition (4.3) useful for describing certain invariants of AFT sofic systems.

This paper was conceived and mostly worked out at the Symposium organized in Summer 1994 at the University of Warwick Maths Institute by Klaus Schmidt and Peter Walters. It is a pleasure to acknowledge the support of the Symposium, and the warmth and stimulation of Warwick's very special mathematical community.

SECTION 1. BACKGROUND AND DEFINITIONS

In this section we assemble some background facts and definitions and establish a little notation. For more details (and proofs) we mention some further references. The book of Lind and Marcus [16] is an excellent introduction to the coding theory of symbolic dynamics. For closing maps, a good place to start is the important paper [14]. The essential reference for finitely presented systems is David Fried's paper [10].

By a (dynamical) system (X, S) (or just S) we will mean a self-homeomorphism S of a compact metrizable space X. Usually, to minimize notation we will use the same name for the homeomorphism and its domain (e.g. referring to a point x in S). A *homomorphism* $\gamma\colon S \to T$ between two systems (X, S) and (Y, T) is a continuous map $\gamma : X \to Y$ such that the obvious diagram commutes:

$$\begin{array}{ccc} X & \xrightarrow{S} & X \\ {\scriptstyle\gamma}\big\downarrow & & {\scriptstyle\gamma}\big\downarrow \\ Y & \xrightarrow{T} & Y \end{array}$$

A homomorphism of systems γ is an *isomorphism* (or topological conjugacy) of systems if it is bijective, a monomorphism (or embedding) if it is injective, and an epimorphism (or *factor map*, or quotient map) if it is surjective. In this latter case we may write $\gamma\colon S \to T$ and refer to T as a quotient or factor of S. Let $\mathcal{F}$ denote the class of factor maps. $\mathcal{F}$ is a category with the operation of composition. STANDING CONVENTION: we always write our compositions from left to right, i.e. $\alpha\beta$ means α followed by β. For any subclass $\mathcal{E}$ of $\mathcal{F}$, a *factorization* of γ in $\mathcal{E}$ is a tuple $(\alpha_1, \ldots, \alpha_k)$ such that each α_i is in $\mathcal{E}$ and $\gamma = \alpha_1\alpha_2 \ldots \alpha_k$. Less formally we will refer to the factorization as just $\gamma = \alpha_1 \ldots \alpha_k$. The factorization is *proper* if every α_i is noninjective. A factor map γ in $\mathcal{E}$ is *irreducible* in $\mathcal{E}$ if γ is not a product of two noninjective factor maps in $\mathcal{E}$. If $\gamma = \alpha\beta$, then we say α is an *initial factor* of γ and β is a *terminal factor* of γ.

If F is a finite set of N symbols, then a point x in the zero dimensional compact metrizable space $F^{\mathbb{Z}}$ is a bisequence $\dots x_{-1}x_0x_1\dots$, where each x_i is in F. This space admits a self homeomorphism S called the N-shift, given by the rule $(Sx)_n = x_{n+1}, n \in \mathbb{Z}$. The restriction of an N-shift to a closed invariant subset is called a subshift. A subshift T is a shift of finite type (SFT) if there is a finite collection of words such that the domain of T is the set of all bisequences (in some N-shift) in which those words never occur.

A *block code* is a homomorphism of subshifts $\gamma\colon\ S \to T$ such that there exists $k \in \mathbb{N}$ and a function g from the set of S-words of length $2k+1$ to the symbol set of T such that for all $x \in S$ and $n \in \mathbb{Z}$, $(\gamma x)_n = g(x_{n-k}\dots x_{n+k})$. The easy but fundamental result of Curtis, Hedlund and Lyndon is that every homomorphism of subshifts is a block code. When $k = 0$, g is just a map on symbols, and γ is called a *one-block code*. If β is a homomorphim of subshifts, then there is an isomorphism α of subshifts such that $\alpha\beta$ is a one-block code. Thus for many purposes we can assume without loss of generality that a factor map between subshifts is a one-block code.

Now suppose G is a finite graph (in this paper, graph always means directed graph) with adjacency matrix A. The associated vertex SFT S_A is the subshift whose symbols are vertices in G and whose domain is the set of bisequences x such that for all n, $A(x_n, x_{n+1}) > 0$. The associated edge SFT S_A is the subshift whose symbols are edges in G and whose domain is the set of bisequences x such that for all n, the terminal vertex of x_n equals the initial vertex of x_{n+1}. Every SFT is isomorphic to a vertex SFT and an edge SFT. If the adjacency matrix is a zero-one matrix, then its vertex SFT and edge SFT are isomorphic. A matrix A is irreducible if it is square and entrywise nonnegative, and for every entry coordinate (i,j) there is some positive n such that $A^n(i,j) > 0$. An SFT is *irreducible* if it is isomorphic to S_A for some irreducible nonnegative integral matrix A. Equivalently, the SFT has a dense forward orbit.

An *irreducible component* of an SFT is a maximal invariant closed subset with a dense forward orbit (equivalently, a maximal invariant closed subset with a dense orbit and dense periodic points). The irreducible components of S_A are in bijective correspondence with the maximal principal irreducible submatrices of A. Associated to an SFT S is its finite graph $\mathcal{G}(S)$ of irreducible components. The vertices of $\mathcal{G}(S)$ are the irreducible components of S. There is at most one edge between two vertices. There is an edge from C to C' iff in S there is a point forwardly asymptotic to C' and backwardly asymptotic to C. (In a system S, we say a point x is forwardly asymptotic, backwardly asymptotic, or biasymptotic to a set L if $\text{dist}(S^n x, L) \to 0$ as $n \to \infty$, $n \to -\infty$, or $|n| \to \infty$ (respectively).)

A *sofic* shift is a subshift which is a quotient of an SFT. Every sofic shift T is a quotient of an SFT by a factor map which is *bounded-to-one* (for some $n \in \mathbb{N}$, every point of T has at most n preimage points). A sofic shift is *irreducible* if it is a quotient of an irreducible SFT; equivalently, it is a quotient of an irreducible SFT by a bounded-to-one factor map; equivalently, it has a dense forward orbit.

A system S is *expansive* if there exists $\epsilon > 0$ such that for all distinct points x, y in S, there exists $n \in \mathbb{Z}$ such that $\text{dist}(S^n x, S^n y) > \epsilon$. This definition does not depend on the choice of compatible metric (although of course the constant ϵ does). A system with zero dimensional domain is isomorphic to a subshift iff it is expansive. A *finitely presented* (FP) system is an expansive quotient of an SFT. A CC system is an FP system with canonical coordinates, i.e. local product structure [10]. The zero dimensional FP systems are the sofic systems, and the zero dimensional CC systems are the SFT's. Krieger [15] showed that certain SFT $\to T$ factor maps are canonically associated to a sofic system T, and Fried generalized this result to show certain CC $\to T$ factor maps are canonically associated to an FP system T [10].

A factor map $\gamma\colon\ S \to T$ is *right closing* if it never collapses distinct backwardly asymptotic points (i.e. if $\gamma x = \gamma y$ and $\lim_{n\to-\infty} \mathrm{dist}\,(S^n x, S^n y) = 0$, then $x = y$). Similarly, γ is left closing if it never collapses distinct forwardly asymptotic points, γ is closing if it is left closing or right closing, and γ is biclosing if it is both left closing and right closing. A bounded-to-one factor map between irreducible SFT's is biclosing iff it is constant-to-one.

There is a notion of a *resolving* factor map between subshifts which is important for constructions [2, 7, 14 ...]. If a factor map β between subshifts is right/left resolving, then it is right/left closing. If it is right/left closing, then there is an isomorphism α of subshifts such that $\alpha\beta$ is right/left resolving. In other words, closing is a topologically invariant formulation of resolving. In some papers (including this one), the identification of isomorphic maps is made tacitly and "resolving" is used as a synonym for "closing".

If $\gamma\colon\ S \to T$ is a factor map betweem irreducible FP systems, then γ satisfies one of the following (mutually exclusive) conditions (in which h denotes the topological entropy).

(1) $h(S) > h(T)$ and γ is generically uncountable-to-one.
(2) $h(S) = h(T)$ and γ is bounded-to-one.

If in addition S is CC, then (2) holds iff γ never collapses distinct biasymptotic points. Define subcategories $\mathcal{B}, \mathcal{B}'$ and $\mathcal{C}$ of $\mathcal{F}$ as follows. For $\gamma \in \mathcal{F}$, $\gamma\colon\ S \to T$,

- $\gamma \in \mathcal{B} \iff \gamma$ is bounded-to-one, S is irreducible SFT and T is sofic.
- $\gamma \in \mathcal{B}' \iff \gamma$ is bounded-to-one, S is irreducible CC and T is FP.
- $\gamma \in \mathcal{C} \iff \gamma$ never collapses distinct biasymptotic points.

Observe, these really are categories (each class is closed under composition of factor maps where compositions are well defined), and $\mathcal{B} \subset \mathcal{B}' \subset \mathcal{C}$. Note, a closing map in particular cannot collapse distinct biasymptotic points. If γ is a factor map from one irreducible FP system onto another, then there is a residual set in the range system all of whose preimage sets have the same cardinality, which is called the *degree* of the factor map.

If $\gamma\colon\ S \to T$ is a factor map, then we let $E(\gamma)$ denote the system which is the restriction of $S \times S$ to $\{(x, y) : \gamma x = \gamma y\}$, which is a closed $(T \times T)$-invariant equivalence relation (which we also call $E(\gamma)$). (Thus, for example, γ is right closing iff in $E(\gamma)$ no point outside the diagonal is backwardly asymptotic to the diagonal.) A closed invariant equivalence relation E gives rise to a quotient system S/E. Two factor maps α and β are *isomorphic* if there are isomorphisms (bijective factor maps) γ and δ such that $\alpha\delta = \gamma\beta$. If two factor maps have the same equivalence relation, then in particular they are isomorphic (with δ above given by the identity map). If S is SFT and $\gamma : S \to T$ is a factor map, then T is expansive (hence FP) iff $E(\gamma)$ is SFT [10, Lemma 1].

Section 2. Finitely many factorizations

Definition 2.1. A subsystem E of a system E' is *convex* in E' if every point of E' biasymptotic to E is in E.

Lemma 2.2. *Suppose that E is convex in an SFT E'.*

(1) *If E intersects an irreducible component C of E', then E contains C.*
(2) *If E intersects irreducible components C_1, C_2 of E', then E contains all orbits backwardly asymptotic to C_1 and forwardly asymptotic to C_2.*
(3) *If E intersects irreducible components C_1, C_3 of E', and in $\mathcal{G}(E(\gamma))$ there are transitions $C_1 \to C_2 \to C_3$, then the component C_2 is also contained in E.*

Proof. Let p be the period of C (the greatest common divisor of the periods of its periodic orbits). The component C is the disjoint union of closed open sets $D_0, \dots, D_{p-1}$ which are cyclically permuted by the shift. Suppose y is a point in D_i. Because E is invariant and intersects C, there is a point x in $E \cap D_i$. There exists M such that for any positive integer n, there is a point w in C such that $w_i = x_i$ when $|i| > n + M$ and $w_i = y_i$ when $|i| \le n$. Because w is biasymptotic to x, it lies in E. It follows that $E \cap C$ is dense in C, and therefore E contains C. This proves (1).

The claim (2) follows from (1). The claim (3) follows because any point of C_2 is a limit of points which are left asymptotic to C_1 and right asymptotic to C_3. $\square$

The lemma explains the following simple proposition.

Proposition 2.3. *Suppose E' is an SFT. If E is convex in E', then E is an SFT, and E is determined by the set of irreducible components of E' which it contains. In particular, there are only finitely many subsystems convex in E'.*

Proposition 2.4. *Suppose $\gamma = \alpha\beta$, where α and β are factor maps. Then $E(\alpha)$ is convex in $E(\gamma)$ iff $\beta \in \mathcal{C}$.*

Proof. Note, $\beta \in \mathcal{C}$ iff the diagonal Δ is convex in $E(\beta)$. Also, $E(\alpha)$ and $E(\gamma)$ are the inverse images under $\alpha \times \alpha$ of Δ and $E(\beta)$. Clearly, if $E(\alpha)$ is not convex in $E(\gamma)$, then Δ is not convex in $E(\beta)$. By a compactness argument the converse also holds. $\square$

Definition 2.5. Suppose γ is a composition of factor maps in two ways, $\gamma = \alpha_1 \dots \alpha_k$ and $\gamma = \alpha'_1 \dots \alpha'_m$ in $\mathcal{C}$. We say that these factorizations of γ are isomorphic if $k = m$ and there exist isomorphisms (topological conjugacies) $\beta_1, \dots, \beta_k$ such that the following diagram commutes:

$$\begin{tikzcd}
{} \arrow[r, "\alpha_1"] \arrow[d, "Id"] & {} \arrow[r, "\alpha_2"] \arrow[d, "\beta_1"] & {} \arrow[r, "\alpha_3"] \arrow[d, "\beta_2"] & \cdots \arrow[r, "\alpha_{k-1}"] & {} \arrow[r, "\alpha_k"] \arrow[d, "\beta_{k-1}"] & {} \arrow[d, "\beta_k"] \\
{} \arrow[r, "\alpha'_1"] & {} \arrow[r, "\alpha'_2"] & {} \arrow[r, "\alpha'_3"] & \cdots \arrow[r, "\alpha'_{k-1}"] & {} \arrow[r, "\alpha'_k"] & {}
\end{tikzcd}$$

Given $k = m$, such a diagram exists iff $E(\alpha_1 \dots \alpha_j) = E(\alpha'_1 \dots \alpha'_j)$ for $1 \le j \le k$.

There is another natural equivalence relation on factorizations, obtained by allowing Id and β_k in the diagram above to be arbitrary isomorphisms. This second relation is a coarser equivalence relation. (For example, if γ and γ' are two factor maps from the 3-shift, each of which collapses two fixed points to a single point and makes no other identifications, then with the coarser relation γ and γ' are isomorphic, whereas with Definition 2.5 they are isomorphic iff they are equal.) Our finiteness results below imply finiteness results for this coarser relation.

Theorem 2.6. *Suppose $\gamma: S \to T$ is a factor map of finitely presented systems. Then up to isomorphism, there are only finitely many factorizations $\gamma = \alpha\beta$ with $\beta \in \mathcal{C}$.*

Proof. Let $\pi: R \to S$ be a factor map such that R is an SFT. A factorization $\gamma = \alpha\beta$ of γ determines a factorization $\pi\gamma = (\pi\alpha)(\beta)$ of $\pi\gamma$, and under this correspondence nonisomorphic factorizations of γ give nonisomorphic factorizations of $\pi\gamma$. So to prove the theorem, without loss of generality we may assume that S is an SFT.

A factorization $\gamma = \alpha\beta$ with $\beta \in \mathcal{C}$ is determined by $E(\alpha)$ which is convex in $E(\gamma)$. Because S is an SFT and T is expansive, $E(\gamma)$ is an SFT, so there can be only finitely many subsystems convex in $E(\gamma)$. $\square$

Theorem 2.7. *Suppose $\gamma: S \to T$ is a factor map of finitely presented systems. Then up to isomorphism, γ has only finitely many proper factorizations in $\mathcal{C}$.*

Proof. As in the previous theorem, without loss of generality we may assume that S is SFT. A proper factorization $\gamma = \alpha_1 \dots \alpha_k$ in $\mathcal{C}$ determines a strictly increasing sequence of SFT's convex in $E(\gamma)$: $\Delta \subsetneq E(\alpha_1) \subsetneq E(\alpha_1\alpha_2) \subsetneq \dots E(\alpha_1 ... \alpha_{k-1}) \subsetneq E(\gamma)$. Two factorizations are isomorphic if and only if they determine the same sequence. There are only finitely many such sequences. $\square$

Corollary 2.8. *Up to isomorphism, an element of $\mathcal{B}$ has only finitely many proper factorizations in $\mathcal{B}$.*

Remarks 2.9.

1. Suppose $\beta: S \to T$ is a bounded-to-one factor map between irreducible FP systems. Let R be an irreducible SFT with $\alpha: R \to S$ bounded-to-one. A proper factorization of β in $\mathcal{C}$ corresponds to a strictly increasing chain of SFT's which are convex in $E(\alpha\beta)$ and contain $E(\alpha)$. Therefore the maximal length of a factorization of β in $\mathcal{C}$ is bounded above by the length of the longest such chain (crudely, by the number of irreducible components of the SFT $E(\alpha\beta)$). This applies in particular to factorizations in $\mathcal{B}$ or $\mathcal{B}'$.

2. Theorem 2.7 implies that for $\mathcal{E} = \mathcal{B}$ or $\mathcal{B}'$, every element in $\mathcal{E}$ is a product of irreducible elements in $\mathcal{E}$. In contrast, every infinite-to-one factor map between irreducible SFT's is a composition of two maps of the same sort [8, Prop. 7.3].

3. Theorem 2.7 also fails for the category of bounded-to-one maps between irreducible sofic systems. It is still possible to make sense of irreducible components of a sofic shift (maximal subsets with dense periodic points and a dense orbit), although they can overlap; the essential problem is that a bounded-to-one map from a sofic shift onto a sofic shift (or onto an SFT) can collapse distinct biasymptotic points. For an example, define a quotient T of a mixing SFT S as follows. Find

disjoint positive-entropy mixing SFT's U, V contained in S such that there is an isomorphism $f\colon\ U \to V$. Let E be the equivalence relation on S generated by $\{(x, fx) : x \in U\}$. Then $E = E(\gamma)$ for a map γ in $\mathcal{B}$ onto an AFT sofic system [6, Lemma 2.4]. For any k we can find a properly nested finite sequence of SFT's in U, $U_1 \subsetneq U_2 \subsetneq \cdots \subsetneq U_k = U$. Let E_i be the equivalence relation on S generated by $\{(x, fx) : x \in U_i\}$. The sequence E_i gives a factorization of γ into k noninjective factor maps between AFT sofic systems.

4. By Theorem 2.6 , up to isomorphism, a factor map γ between FP systems has only finitely many factorizations $\gamma = \alpha\beta$ with β a closing map. The previous example shows that γ can have infinitely many nonisomorphic factorizations $\gamma = \alpha\beta$ where α is a closing map between FP systems.

As a matter of general symbolic dynamics, one would like to understand which dynamical systems are quotients of SFT's. Below is an application of our viewpoint to this problem. First we recall the definition of an inverse limit of dynamical systems. We are given systems $T_n\colon\ X_n \to X_n,\ n \geq 1$, and homomorphisms ("bonding maps") $\pi_n\colon T_n \to T_{n-1}$. The resulting inverse limit system is defined to be the restriction of the product system $\prod T_n$ to the subset of $\prod X_n$ consisting of the sequences $(x_1, x_2, \dots)$ such that $\pi_n(x_n) = x_{n-1}$ for every n.

Theorem 2.10. *Suppose T is the inverse limit of irreducible CC systems $T_n, 1 \leq n < \infty$, by surjective, bounded-to-one, noninjective bonding maps $\pi_n : T_n \to T_{n-1}$. Then T is not a quotient of an SFT.*

Proof. Suppose there is a factor map γ from some SFT S onto T. Let γ_n be the factor map defined by following γ with the projection of T onto T_n. This gives a system of factor maps $\gamma_n\colon\ S \to T_n$ such that $\gamma_n = \gamma_{n+1}\pi_{n+1}$, so $\gamma_1 = (\gamma_n)(\pi_n\pi_{n-1}\dots\pi_2)$ if $n \geq 2$. But the factor maps $(\pi_n\pi_{n-1}\dots\pi_2)$ are all in $\mathcal{C}$, and this contradicts Theorem 2.6. $\square$

Example 2.11. Define an inverse limit system T by letting each T_n equal the 2-shift, and taking for each bonding map the standard two-to-one map $x \mapsto y$ given by $y_i = x_i + x_{i+1} (\mathrm{mod}\ 2)$. Despite its simple algebraic structure, T is not a quotient of a shift of finite type.

Remark 2.12. We mention one more application: if $\beta\colon\ S \to T$ is a factor map in $\mathcal{C}$, T is finitely presented, and S is not expansive, then S cannot be the quotient of an SFT. For if R is an SFT and $\alpha\colon\ R \to S$ is a factor map, then (by Prop. 2.4) $E(\alpha)$ is convex in the SFT $E(\alpha\beta)$, so (by Prop. 2.3) $E(\alpha)$ is SFT and therefore $S/E(\alpha)$ is expansive.

Section 3. Left/right factorizations

By a *left/right factorization* of a factor map γ we will mean a proper factorization $\gamma = \alpha\beta$ where α is left closing and β is right closing. (Of course all our statements about left/right factorizations can be translated to statements about right/left factorizations.) The only known general methods for constructing bounded-to-one factor maps between irreducible SFT's and sofic systems involve closing maps and their compositions, but such maps do constitute a rich and useful class (e.g., [2, 3, 7, 14]). The existence of left/right factorizations for a given factor map $\gamma \in \mathcal{B}$ was considered in [14] and especially in [21]. In this section, we'll simplify and extend results of Trow [21] using equivalence relations.

Definition 3.1. A subsystem E of a system E' is *forward-closed* in E' if whenever a point x of E' is backwardly asymptotic to E, then $x \in E$.

Note, if E is forward-closed in E', then E is convex in E'.

Lemma 3.2. *Suppose $\gamma = \alpha\beta$ is a factorization of factor maps.*

(1) *α is left closing iff $E(\alpha) - \Delta$ contains no point forwardly asymptotic to Δ.*
(2) *β is right closing iff $E(\alpha)$ is forward closed in $E(\gamma)$.*

Proof. (1) is clear. For (2), note that β is right closing iff no point of $E(\beta) - \Delta$ is backwardly asymptotic to Δ. Taking inverse images under α and applying compactness, we see this holds iff no point of $E(\gamma) - E(\alpha)$ is backwardly asymptotic to $E(\alpha)$, i.e., iff $E(\alpha)$ is forward closed in $E(\gamma)$. $\square$

Proposition 3.3. *Suppose $\gamma \in \mathcal{B}$, $\gamma : S \to T$, and E is a closed $(S \times S)$-invariant equivalence relation contained in $E(\gamma)$. Then the following are equivalent.*

(1) *There are factor maps α, β between sofic shifts such that $\gamma = \alpha\beta$, α is left closing, β is right closing, and $E(\alpha) = E$.*
(2) *E is forward-closed in $E(\gamma)$, and contains no point outside the diagonal which is forwardly asymptotic to the diagonal.*

Proof. Given the lemma, it remains to check, for a factorization $S \to S/E \to T$ given by E from (2), that S/E is sofic. Since E is forward closed in $E(\gamma)$, it is convex in $E(\gamma)$. Therefore E is SFT, so the quotient system S/E is expansive. As a bounded-to-one extension of the zero-dimensional system T, it must also be zero dimensional [13, Thm. VI 7]. Therefore S/E is sofic. $\square$

Example 3.4. For concreteness, we describe a factor map $\gamma \colon S \to T$ between irreducible SFT's such that γ has a right/left factorization in $\mathcal{B}$, but γ has no left/right factorization. (In [14], there is a factor map $\gamma \colon S \to T$ between irreducible SFT's, where γ has a right/left factorization in $\mathcal{B}$, but no (!) factor map $S \to T$ has a left/right factorization in $\mathcal{B}$.) Also there are E convex in $E(\gamma)$ such that S/E is strictly sofic (i.e., sofic but not SFT).

Let S be a vertex SFT defined by the graph on vertices $\{a, b, c_1, c_2, c_3, d, e\}$ with edges $a \to c_1, c_2$; $b \to c_2, c_3$; $c_1 \to c_1, e$; $c_2 \to c_2, d$; $c_3 \to c_3, e$; $d \to e$; $e \to a, b$. Define $\gamma \colon S \to T$ as the one-block code which deletes subscripts. The initial (right-closing) factor for the right/left factorization of γ is the one-block code which identifies only c_1 and c_3. The only other factor maps with equivalence relations convex in $E(\gamma)$ are given by the one-block code which identifies only c_1 and c_2 and the one-block code which identifies only c_2 and c_3. For each of these, S/E is strictly sofic.

Definition 3.5. Given a factor map γ, let $E_F(\gamma)$ be the smallest closed invariant equivalence relation in $E(\gamma)$ which is forward-closed in $E(\gamma)$.

Proposition 3.6. *Suppose $\gamma: S \to T$ is a bounded-to-one factor map of irreducible sofic systems. Then γ has only finitely many left/right factorizations (up to isomorphism). When S is SFT, the following are equivalent.*

(1) *γ has a left/right factorization.*
(2) *$S \to S/E_F(\gamma) \to T$ is a left/right factorization.*
(3) *$E_F(\gamma) - \Delta_S$ has no points forwardly asymptotic to Δ_S.*

Proof. For S SFT, the finiteness claim follows from Propositions 2.3 and 3.3 (recall, if $E(\alpha)$ is forward-closed in $E(\gamma)$, then $E(\alpha)$ is convex in $E(\gamma)$). The finiteness claim for S sofic is a corollary, by considering the precomposition of γ with a left closing factor map from an irreducible SFT. The equivalence of (2) and (3) follows from Lemma 3.2, and it is trivial that (2) implies (1). If $\gamma = \alpha\beta$ is a left/right factorization, then by Prop. 3.3, $E(\alpha)$ is forward-closed, and thus contains $E_F(\gamma)$. By Proposition 3.3, $E(\alpha) - \Delta_S$ has no points forwardly asymptotic to Δ_S, and therefore neither does $E_F(\gamma) - \Delta_S$. $\square$

Trow [21] proved that a map γ in $\mathcal{B}$ has only finitely many left/right factorizations in $\mathcal{B}$, and if degree$(\gamma) = 1$ then γ has at most one such factorization. We can now recover his result.

Theorem 3.7. *Suppose $\gamma \in \mathcal{B}$ has a left/right factorization $S \to S/E \to T$ such that S/E is SFT. Then $S/E_F(\gamma)$ is SFT and the natural factor map $p : S/E_F(\gamma) \to S/E$ is constant-to-one. In particular, if γ has degree 1, then $p = Id$ and the left/right factorization of γ is unique (up to isomorphism).*

Proof. Necessarily, $E_F(\gamma) \subset E$. We provide some names for our commuting diagram of factor maps:

$$\begin{array}{ccccc} S & \xrightarrow{\ l'\ } & S/E_F(\gamma) & \xrightarrow{\ r'\ } & T \\ \big\downarrow{\scriptstyle Id} & & \big\downarrow{\scriptstyle p} & & \big\downarrow{\scriptstyle Id} \\ S & \xrightarrow{\ l\ } & S/E & \xrightarrow{\ r\ } & T \end{array}$$

The maps l, l' are left closing and the maps r, r' are right closing. Let E' denote $E_F(\gamma)$.

The map p must be right closing because it is an initial factor of the right closing map r'. Because S/E' is irreducible and S/E is SFT and there is a right closing map (p) from S/E' onto S/E, it follows that S/E' is SFT [7, Prop. 4.12] Because a left closing map (l) between irreducible sofic systems has p has a terminal factor *and* p is a factor map between SFT's, it follows that p is left closing [7, Props. 4.10 and 4.11]. Now p is a biclosing map between two irreducible SFT's, and therefore is constant-to-one [18, Thm. 7.3; 5, Prop. 1]. $\square$

Example 3.8. The uniqueness claim in the last theorem can fail when $S/E_F(\gamma)$ is not SFT. For example, let S be the full shift on N symbols and define $\gamma: S \to T$ to be the map collapsing all fixed points of S to a single point, and making no other identifications. Then T is a "near Markov" [6] sofic shift and any equivalence relation contained in $E(\gamma)$ is the equivalence relation for an initial factor of a left/right factorization of γ.

Example 3.9. For a degree 1 factor map $\gamma\colon S \to T$ in $\mathcal{B}$, it can happen that γ has two left/right factorizations $S \to S/E_F(\gamma) \to T$ and $S \to S/E \to T$ where the induced map $p : S/E_F(\gamma) \to S/E$ is not left closing (p is always right closing). (By Theorem 3.7, in this case S/E and $S/E_F(\gamma)$ must be strictly sofic.) For completeness we give an example. Define S as a vertex SFT from a graph with vertex set $\{a_1, a_2, b_1, b_2, c, d, e, f_1, f_2\}$. The map γ will be the one-block code which erases subscripts. In the graph, each subscripted vertex has an edge to itself, and the remaining transitions are $f_1 \to c \to a_1 \to d \to f_1$; $f_2 \to c \to a_2 \to e \to f_2$; $d \to b_1 \to a_1$; and $e \to b_2 \to a_2$. The identification $a_1 = a_2$ gives $S \to S/E_F(\gamma)$. The identifications $a_1 = a_2$, $b_1 = b_2$ give $S \to S/E$.

Section 4. Decision procedures

Factorizations in $\mathcal{B}$ (4.1).

Given $\gamma \in \mathcal{B}$, we will indicate how to list up to isomorphism all proper factorizations of γ in $\mathcal{B}$. Then we will point out how to determine which maps in these factorizations are closing maps.

A proper factorization $\gamma = \alpha_1 \dots \alpha_k$ in $\mathcal{B}$ determines equivalence relations $E_i = E(\alpha_1 \dots \alpha_i)$, $1 \le i < k$, such that each E_i is convex in $E(\gamma)$, each S/E_i is SFT, $\Delta \subsetneq E_i \subsetneq E(\gamma)$ for each i, and $E_i \subsetneq E_{i+1}$ if $1 \le i < k-1$. Conversely, such a sequence of equivalence relations $E_1, \dots, E_{k-1}$ determines the isomorphism class of a proper factorization of γ in $\mathcal{B}$. By the convexity (Lemma 2.2), any such equivalence relation E_i is determined by the set $K(E_i)$ of its irreducible components, which form a proper subset of the irreducible components of the SFT $E(\gamma)$. Thus the increasing sequences $E_1 \subsetneq \cdots \subsetneq E_{k-1}$ are determined by the increasing sequences $K(E_1) \subsetneq \cdots \subsetneq K(E_{k-1})$ of their sets of irreducible components. Because we can enumerate the finite collection of proper subsets of the irreducible components of $E(\gamma)$, it follows that we can list the proper factorizations of γ in $\mathcal{B}$ if we have a decision procedure for the following problem: given a proper subset K of the irreducible components of $E(\gamma)$, determine whether K is the set of irreducible components of an equivalence relation E such that E is convex in $E(\gamma)$ and S/E is an SFT. (If there is no such K, then γ has no proper factorization in $\mathcal{B}$.) It remains to describe a decision procedure for this problem.

For the given K, let E_K be the subset of $E(\gamma)$ consisting of all points forwardly and backwardly asymptotic to components in K. If E_K contains any irreducible component of $E(\gamma)$ which is not in K, then K cannot be the set of irreducible components of a convex equivalence relation in $E(\gamma)$, and we are done. So suppose that K does equal $K(E_K)$. Then E_K will be an equivalence relation convex in $E(\gamma)$ if E_K is an equivalence relation.

It is trivial to check whether E_K is reflexive and symmetric; if it is not, then we have determined that K is not a set of the required form; so suppose that E_K is reflexive and symmetric. $E(\gamma)$ is contained in the product $S \times S$, and the symbols of the SFT $E(\gamma)$ are pairs (a, b) of symbols from S. Without loss of generality we may suppose that $E(\gamma)$ is a one-step SFT: a point (x, y) lies in the SFT $E(\gamma)$ iff for every n the words $(x_n, y_n)(x_{n+1} y_{n+1})$ occur in $E(\gamma)$ (this holds if γ is given as a one-block code). Then E_K is also a one-step SFT. Form a new one-step SFT R whose symbols are triples (a, b, c) of symbols from S, where a word $(a, b, c)(a', b', c')$ is allowed iff

$(a, b)(a', b')$ and $(b, c)(b', c')$ are allowed words in E_K. Now, the relation E_K is not transitive iff there are points x, y, z in S such that (x, y) and (y, z) are in E_K but (x, z) is not. This happens iff the image of R under the map $(x, y, z) \mapsto (x, z)$ contains a point outside E_K. This is straightforward to check. Therefore we can determine whether E_K is transitive. If it is not, then we are done; so suppose that E_K is transitive, i.e. E_K is an equivalence relation convex in $E(\gamma)$.

Even when E_K is an equivalence relation convex in $E(\gamma)$, it can happen that S/E_K is strictly sofic (see Example 3.4). However, Coven and Paul [9] gave an algorithm for checking whether a factor map from an SFT onto a subshift has image which is SFT. We appeal to their algorithm to conclude our decision procedure.

Finally, given a factorization $\gamma = \alpha_1 \dots \alpha_k$ in $\mathcal{B}$, we can easily determine which α_i are closing maps. By Lemma 3.2, α_i is right closing iff $E(\alpha_1 \dots \alpha_{i-1})$ is forward closed in $E(\alpha_1 \dots \alpha_{i-1}\alpha_i)$ (where $E(\alpha_1 \dots \alpha_{i-1})$ is interpreted as the diagonal if $i = 1$). Here $E(\gamma)$ is SFT and this condition can be easily checked from a graph presenting $E(\gamma)$. A similar procedure (with "backward closed" replacing "forward closed") works to check left closing.

In [21] Trow gave a procedure for determining, given $\gamma \in \mathcal{B}$, whether there exists a left/right factorization $\gamma = \alpha\beta$ in $\mathcal{B}$ for which β has degree 1. In [22] he gave a procedure for determining the factorizations in $\mathcal{B}$ for which β is biclosing.

Factorizations in $\mathcal{B}'$ (4.2).

Let $\mathcal{B}'$ be the category of bounded-to-one factor maps from irreducible CC systems to FP systems. Suppose $\beta \in \mathcal{B}'$, $\beta\colon\ S \to T$. By Theorem 2.7, up to isomorphism β has only finitely many factorizations in $\mathcal{B}'$. Let $\alpha\colon\ R \to S$ be a bounded-to-one factor map from an irreducible SFT. Up to isomorphism, the factorizations of β in $\mathcal{B}'$ are in bijective correspondence with the strictly increasing chains of closed invariant equivalence relations E convex in $E(\alpha\beta)$ such that E contains $E(\alpha)$ and S/E is CC. The collection of such relations E can be determined as in the procedure for $\mathcal{B}$, except that another procedure is required to determine whether S/E is CC. Nicolai Haydn was the first to offer such a procedure [11].

There is probably no really simple algorithm for determining whether an FP image of a CC system is CC (in the zero-dimensional case, this is asking when a subshift image of an SFT is SFT). In general this seems to be a more or less accidental matter of whether images of stable and unstable sets overlap in a way satisfying the Markov condition.

The equivalence relation viewpoint gives the following easy answer to a problem which arises in studying AFT sofic systems [6].

Proposition 4.3. *Suppose S' is a subshift of an SFT S and E' is an equivalence relation on S' such that $S' \to S'/E'$ is a biclosing map onto a subshift. Let $E = E' \cup \Delta_S$. Then S/E is sofic iff $E' - \Delta_{S'}$ is SFT.*

Proof. S/E is zero dimensional because it is the union of the closed zero dimensional set S'/E' and the zero dimensional set $S/E - S'/E'$ (which is homeomorphic to the zero dimensional set $S - S'$) [13, Cor. 2 of Thm. II 2]. Thus S/E will be sofic iff it is expansive iff E is SFT. Because $S' \to S'/E'$ is biclosing, $\Delta_{S'}$ is an isolated subset of E', so $E' - \Delta_{S'}$ is a closed set. Because Δ_S is SFT and E is the disjoint union of the closed invariant sets Δ_S and $E' - \Delta_{S'}$, it follows that E is SFT iff $E' - \Delta_{S'}$ is SFT. $\square$

Given an irreducible SFT S, the proposition can be used with Krieger's Embedding Theorem to characterize the factor maps α such that there exists an AFT sofic system T and a biclosing degree 1 factor map $\pi\colon S \to T$ whose restriction to $\{x : |\pi^{-1}(\pi(x))| > 1\}$ is isomorphic to α (see [6]). Such a cover π is canonically associated to T [15, 5], so the isomorphism classes of S (as a dynamical system) and α (as a factor map) are invariants of the dynamical system T.

The hypothesis in the proposition that S'/E' is a subshift can be removed. This is because an FP system which is the image of an SFT under a biclosing map must be zero-dimensional. We forego a proof here to avoid a digression into the topology of FP systems.

References

1. R. Adler, D. Coppersmith & M. Hassner, *Algorithms for sliding block codes*, IEEE Trans. Info.Th., Vol. IT-29 (1983), 5-22.
2. R. Adler & B. Marcus, *Topological entropy and equivalence of dynamical systems*, Memoirs AMS 219 (1979).
3. J. Ashley, *Resolving factor maps for shifts of finite type with equal entropy*, Ergod. Th. & Dynam. Sys. 11 (1991), 219-240.
4. M. Boyle, *Symbolic dynamics and matrices*, Combinatorial and graph-theoretical problems in linear algebra (R. A. Brualdi, S. Friedland & V. Klee, eds.), I.M.A. Volumes in Math. 50, Springer-Verlag, pp. 1-38.
5. M. Boyle, B. Kitchens & B. Marcus, *A note on minimal covers for sofic systems*, Proc. AMS 95 (1985), 403-411.
6. M. Boyle & W. Krieger, *Almost Markov and shift equivalent sofic systems*, Dynamical Systems, Proceedings of the Special Year in Dynamics 1986-87 at the University of Maryland (J.C. Alexander, ed.), Springer Lecture Notes in Math 1342, pp. 33-93.
7. M. Boyle, B. Marcus & P. Trow, *Resolving Maps and the Dimension Group for Shifts of Finite Type*, Memoirs AMS 377 (1987).
8. M. Boyle & S. Tuncel, *Infinite-to-one Codes and Markov measures*, Trans. AMS 285, No. 2 (1984), 657-684.
9. E. Coven & M. Paul, *Finite procedures for sofic systems*, Monats. fur Math. 83 (1977), 165-177.
10. D. Fried, *Finitely presented dynamical systems*, Ergod. Th. & Dynam. Sys. 7 (1987), 489-507.
11. N.T.A. Haydn, *The product structure of finitely presented dynamical systems*, Israel J. Math. 73, No. 3 (1991), 343-356.
12. C. Heegard, B. Marcus & P. Siegel, *Variable length state splitting with applications to average run-length constrained (ARC) codes*, IEEE-Information Theory 37 (1991), 759-777.
13. W. Hurewicz & H. Wallman, *Dimension Theory*, Princeton University Press, Princeton NJ, 1948.
14. B. Kitchens, B. Marcus & P. Trow, *Eventual factor maps and compositions of closing maps*, Ergod. Th. & Dynam. Sys. 11 (1991), 85-113.
15. W. Krieger, *On sofic systems I*, Israel J. Math 48 (1984), 305-330.
16. D. Lind & B. Marcus, *An introduction to symbolic dynamics and coding*, Cambridge University Press (1995).
17. B. Marcus, *Sofic systems and encoding data*, IEEE-IT 31 (1985), 366-377.
18. M. Nasu, *Constant-to-one and onto global maps of homomorphisms between strongly connected graphs*, Ergod. Th. & Dynam. Syst. 3 (1983), 387-413.
19. M. Nasu, *Topological conjugacy for sofic systems and extensions of automorphisms of finite subsystems of topological Markov shifts*, Proceedings of Maryland Special Year in Dynamics 1986-87. Springer Lecture Notes in Math 1342 (1988), Springer-Verlag.

20. W. Parry, *Notes on coding problems for finite state processes*, Bull. London Math. Soc. 23 (1991), 1-33.
21. P. Trow, *Decompositions of finite-to-one factor maps*, Israel J. Math. 91 (1995), 129-155.
22. P. Trow, *Decompositions involving bi-closing maps*, preprint (1994).

Dept. of Mathematics, Univ. of Maryland, College Park MD 20742 U.S.A.
E-mail address: mmb@math.umd.edu