Invertible Binary Transducers and Automorphisms of the Binary Tree

Tsutomu Okano

May 5, 2015

Abstract

An invertible binary transducer is a Mealy machine with input and output alphabet $2 = \{0, 1\}$ such that the transition function at each state is a permutation of $\{0, 1\}$. Each state may be interpreted as a start state, and the transduction is an automorphism of 2^*. These transductions generate a subsemigroup $\mathcal{S}(\mathcal{A})$ and a subgroup $\mathcal{G}(\mathcal{A})$ of $Aut(2^*)$. In this research we focus on the instances when the semigroup is commutative. In this case, Nekrashevych and Sidki have shown that $\mathcal{G}(\mathcal{A})$ must be either a finite 2-group or a finite dimensional free abelian group via a linear algebraic approach. We intend to gain more understanding about the structure of these finite state machines, and explore questions on the rationality of relations given by orbits induced by automorphisms from $\mathcal{S}(\mathcal{A})$.

Contents

1 Introduction

A finite state transducer is a 5-tuple $(Q, s_0, \Sigma, \Gamma, \delta)$ where Q is a finite set of states with a designated start state $s_0 \in Q$, Σ and Γ are finite alphabets, and

$$\delta : Q \times \Sigma \to Q \times \Gamma$$

is a transition function. A transition is usually labeled as

$$s \xrightarrow{a/b} t$$

where s and t are states and $a \in \Sigma$ and $b \in \Gamma$. A transducer is said to be invertible if $\Sigma = \Gamma$ with size n and each state $s \in Q$ has a permutation $\pi_s \in S_n$ such that $\delta(s, a) = (s_a, \pi_s(a))$. If π_s is the identity permutation, s is said to be a copy-state. If π_s is the non-identity permutation, s is said to be a toggle state. The transduction preserves the structure of the infinite n-ary tree, and thus belongs to the automorphism group of the tree. This paper will explore the most basic case when $n = 2$. The full binary tree will be denoted as 2^*, and its automorphism group will be denoted as $Aut(2^*)$.
Given an invertible binary transducer $\mathcal{A}$, another invertible binary transducer with the same transition structure by choosing some other state $s \in Q$ to be the start state. For a state s, let $\underline{s}$ denote the transduction given by the associated transducer. Then $\mathcal{A}$ uniquely determines a finitely generated subsemigroup

$$\mathcal{S}(\mathcal{A}) = \langle \underline{s} : s \in Q \rangle$$

of $Aut(2^*)$. Further, the subgroup $\mathcal{G}(\mathcal{A})$ of $Aut(2^*)$ generated by these automorphisms have recently gained popularity through application in group theory. A notable example is Grigorchuk's 5-state machine, pictured below.

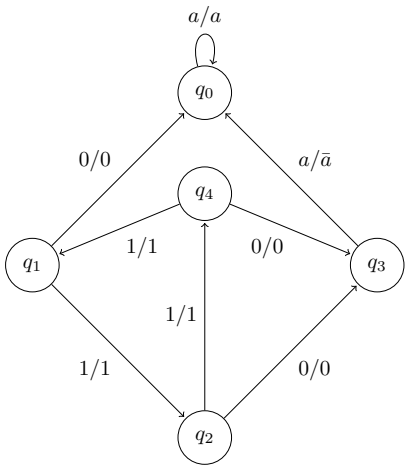

Figure 1: The Grigorchuk machine

Here, the subgroup generated by the five automorphisms (note that $\underline{q_0}$ is just the identity function) is of intermediate growth and was the first constructive answer to Milnor's question of whether such a group exists [1]. Another important example is Aleshin's machine. This gives a simple example for Burnside's problem which asks whether there exists a finitely generated infinite group such that every element has finite order [2].
As these examples indicate, the properties of $\mathcal{S}(\mathcal{A})$ and $\mathcal{G}(\mathcal{A})$ are complicated in general. However, this paper restricts the attention to the commutative cases. A fundamental result by Nekrashevych and Sidki is that in such a case, $\mathcal{G}(\mathcal{A})$ is either isomorphic to a finite Boolean group or to $\mathbb{Z}^m$ for some $m \geq 1$, and the transitions in $\mathcal{A}$ can be encoded as a pair (A, r) of a matrix and a vector [3]. This theorem will be stated

in Section 3 asTheorem 3. This connection will be further explained in Section 3. This paper answers some decidability questions raised by Sutner and Lewi [4] and expand upon the linear algebraic methods employed by Nekrashevych and Sidki.

2 Commutative Semigroups

This section introduces commutative transducers and gives basic structure results for such transducers and the associated semigroup $\mathcal{S}(\mathcal{A})$. The notion of residue function ∂ and its associated calculus is explored. Finally, a polynomial time algorithm for determining whether $\mathcal{S}(\mathcal{A})$ is commutative is presented.

Definition 2.1. An invertible binary transducers $\mathcal{A}$ is said to be *commutative* or *abelian* iff $\mathcal{S}(\mathcal{A})$ is commutative.

Definition 2.2. Two states s and t in a transducer $\mathcal{A}$ are said to *commute* iff $\underline{s}$ and $\underline{t}$ commute as elements of $Aut(2^*)$.

Definition 2.3. A transducer $\mathcal{A}$ is said to be *minimal* if each state encodes a unique automorphism. That is, there do not exist two distinct states s, s' with $\underline{s} = \underline{s'}$.

Regard $\mathcal{A}$ as a labeled directed graph and consider the subgraph G of $\mathcal{A}$ obtained by keeping all the states and removing all out-edges from the toggle states. The following lemma provides a necessary condition for $\mathcal{A}$ to be commutative.

Lemma 2.1. *Suppose that $\mathcal{A}$ is commutative and minimal. Then each connected component of G is either a directed path leading to a toggle state, or a self-loop on a copy state. In particular, there exists at most one self-looping copy state.*

Proof. By commutativity of $\mathcal{A}$, the two transitions of a copy state must coincide (shown by Nekrashevych and Sidki).
If a self-looping copy state exists, its uniqueness is a consequence of minimality. Minimality also guarantees that the self-looping copy state must be isolated in G.
Let C be a connected component of G that is not just a self-looping copy state. By minimality, this implies that C contains no directed cycle of copy states so C must be a DAG. Further, since each copy state has a unique copy state to transition to in C, it must be a rooted tree. In particular, the tree is directed toward a unique toggle state, since a toggle state does not have an out-edge in G. If C is not a path, there exist two distinct states a and b in C such that their distances from the root toggle state are equal and minimal. Since they both transition into the same state, it violates the minimality condition. Hence, C is a directed path terminating at a toggle state. □

Definition 2.4. For a toggle state t , the directed path of copy states leading to t, if it exists, will be referred to as the *copy chain of t*.

Definition 2.5. For $x \in 2^*$ and $f \in Aut(2^*)$, the *residual of f by x* is $\partial_x f \in Aut(2^*)$ defined so that, for $y \in 2^*$, $\partial_x f(y)$ is the suffix of $f(x :: y)$ obtained by removing $f(x)$.

The following lemma controls the behavior of transitions from toggle states.

Lemma 2.2. *Suppose that $\mathcal{A}$ is commutative. Then there exists a unique $\theta_{\mathcal{A}} \in \mathcal{G}(\mathcal{A})$ such that for every toggle state t,*

$$\partial_1 \underline{t} = (\partial_0 \underline{t})\theta_A$$

is satisfied.

Proof. Let t_i, t_j be two distinct toggle states. By evaluating ∂_0 on both sides of

$$\underline{t_i}\underline{t_j} = \underline{t_j}\underline{t_i}$$

the following is obtained,

$$(\partial_0 \underline{t_i})(\partial_1 \underline{t_j}) = (\partial_0 \underline{t_j})(\partial_1 \underline{t_i})$$

Since $\mathcal{G}(\mathcal{A})$ is commutative, it follows that

$$(\partial_0 \underline{t_i})^{-1}\partial_1 \underline{t_i} = (\partial_0 \underline{t_j})^{-1}\partial_1 \underline{t_j}$$

t_i and t_j were chosen arbitrarily, so there exists a unique $\theta_{\mathcal{A}}$ such that

$$\theta_{\mathcal{A}} = (\partial_0 \underline{t})^{-1}\partial_1 \underline{t} \in \mathcal{G}(\mathcal{A})$$

holds true for every toggle state t. □

Definition 2.6. The function $\theta_{\mathcal{A}}$ defined as above will be referred to as the *shift function* of $\mathcal{A}$.

Corollary 2.1. *Suppose that $\mathcal{A}$ is commutative and suppose that there is a toggle state such that the two residuals are equal. Then for any toggle state, the two residuals are equal. Furthermore, every state has itself as the inverse.*

Proof. The shift function is the identity function, so the first result holds immediately. The second result is an easy corollary. □

It follows that $\mathcal{S}(\mathcal{A}) = \mathcal{G}(\mathcal{A})$ and by minimality, $\mathcal{G}(\mathcal{A})$ is isomorphic to the Boolean group $(\mathbb{Z}/2\mathbb{Z})^n$, where n is the number of states in $\mathcal{A}$.

Corollary 2.2. *Suppose that $\mathcal{A}$ is commutative and minimal. Then for every pair of distinct toggle states t_1 and t_2,*

$$\partial_0 \underline{t_1} \neq \partial_0 \underline{t_2}$$

and

$$\partial_1 \underline{t_1} \neq \partial_1 \underline{t_2}$$

In other words, residuation by 0 and residuation by 1 are injections for all states.

Proof. If $\partial_0 \underline{t_1} = \partial_0 \underline{t_2}$, then

$$\partial_1 \underline{t_1} = (\partial_0 \underline{t_1})\theta_{\mathcal{A}} = (\partial_0 \underline{t_2})\theta_{\mathcal{A}} = \partial_1 \underline{t_2}$$

which contradicts the minimality. The other case is proved in a symmetric fashion. □

Corollary 2.3. *Exponentiation rules for $f \in \mathcal{G}(\mathcal{A})$ if $\mathcal{A}$ is commutative :*

1. *If f is even,*
$$f^n = ((\partial_0 f)^n, (\partial_0 f)^n)$$
2. *If f is odd,*
$$f^n = \begin{cases} ((\partial_0 f)^n \theta_{\mathcal{A}}^k, (\partial_0 f)^n \theta_{\mathcal{A}}^k) & \text{if } n = 2k \\ ((\partial_0 f)^n \theta_{\mathcal{A}}^{k+1}, (\partial_0 f)^n \theta_{\mathcal{A}}^k)\sigma & \text{if } n = 2k+1 \end{cases}$$

Proof. By commutativity, it suffices to prove the claim for the generators. This can be done for each generator by induction on the exponent in both directions. □

Note that $\theta_{\mathcal{A}}$ still satisfies

$$\partial_1 f = (\partial_0 f)\theta_{\mathcal{A}}$$

whenever $f \in \mathcal{G}(\mathcal{A})$ is an odd function. The following theorem characterizes a condition for $\mathcal{S}(\mathcal{A})$ to be commutative.

Theorem 1. *Suppose that $\mathcal{A}$ is minimal. Then $\mathcal{A}$ is commutative iff there is a unique choice of $\theta_{\mathcal{A}}$ such that each state t in $\mathcal{A}$ satisfies*

$$\partial_1 \underline{t} = \begin{cases} \partial_0 \underline{t} & \text{if } t \text{ is a copy state} \\ (\partial_0 \underline{t})\theta_{\mathcal{A}} & \text{if } t \text{ is a toggle state} \end{cases}$$

Proof. Left implies right is a consequence of previous lemmas. Right implies left is proven by showing that any two states act commutatively on a string of length at most k, by induction on k.
Base case : $k = 1$ is trivial.
Induction step : Assume that the claim holds true for some $k \geq 1$. Let x be a string of length $k+1$, where without loss of generality $x = 0 : y$ for some $y \in 2^k$. Let t_1, t_2 be two states in $\mathcal{A}$. There are three cases to examine, all invoking the induction hypotheses on the residuals.

1. t_1 and t_2 are copy states :

$$\begin{aligned} x\underline{t_1 t_2} &= \big(0 : (y\ \partial_0 \underline{t_1})\big)\ \underline{t_2} \\ &= 0 : \big(y\ (\partial_0 \underline{t_1})(\partial_0 \underline{t_2})\big) \\ &= 0 : \big(y\ (\partial_0 \underline{t_2})(\partial_0 \underline{t_1})\big) \\ &= \big(0 : (y\ \partial_0 \underline{t_2})\big)\ \underline{t_1} \\ &= x\underline{t_2 t_1} \end{aligned}$$

2. t_1 is a copy state and t_2 is a toggle state :

$$\begin{aligned} x\underline{t_1 t_2} &= \big(0 : (y\ (\partial_0 \underline{t_1}))\big)\ \underline{t_2} \\ &= 1 : \big(y\ (\partial_0 \underline{t_1})(\partial_0 \underline{t_2})\big) \\ &= 1 : \big(y\ (\partial_0 \underline{t_2})(\partial_0 \underline{t_1})\big) \\ &= \big(1 : (y\ (\partial_0 \underline{t_2}))\big)\ \underline{t_1} \\ &= x\underline{t_2 t_1} \end{aligned}$$

3. t_1 and t_2 are toggle states :

$$\begin{aligned} x\underline{t_1 t_2} &= \big(1 : (y(\partial_0 \underline{t_1}))\big)\ \underline{t_2} \\ &= 0 : \big(y(\partial_0 \underline{t_1} \partial_1 \underline{t_2})\big) \\ &= 0 : \big(y(\partial_0 \underline{t_1} \partial_0 \underline{t_2} \theta_{\mathcal{A}})\big) \\ &= 0 : \big(y(\partial_0 \underline{t_2} \partial_0 \underline{t_1} \theta_{\mathcal{A}})\big) \\ &= 0 : \big(y(\partial_0 \underline{t_2} \partial_1 \underline{t_1})\big) \\ &= \big(1 : (y(\partial_0 \underline{t_2}))\big)\ \underline{t_1} \\ &= x\underline{t_2 t_1} \end{aligned}$$

By induction, any pair of states in $\mathcal{A}$ act commutatively on any string in 2^*. □

Lemma 2.3. *For any transducer $\mathcal{A}$, there exists a transducer whose states represent exactly the inverses of automorphisms given by states in $\mathcal{A}$.*

Proof. Let $s_1, ..., s_n$ be states of $\mathcal{A}$. Consider a transducer $\mathcal{A}'$ with states $t_1, ..., t_n$ with transitions defined as

$$\begin{aligned} \partial_0 \underline{t_i} &= \partial_1 \underline{s_i} \\ \partial_1 \underline{t_i} &= \partial_0 \underline{s_i} \end{aligned}$$

and t_i is a toggle state if and only if s_i is a toggle state. Then it can be shown that $\underline{s_i}^{-1} = \underline{t_i}$ by induction on string lengths. □

Definition 2.7. It is well known that the transducer built above is referred to as the *inverse transducer* of $\mathcal{A}$ and is denoted as $\mathcal{A}^{-1}$.

The inverse relation is well-defined in the sense that it is an involution.

Definition 2.8. For two binary strings of the same length $x = x_0...x_n$ and $y = y_0...y_n$, the *convolution* of x and y is defined to be the string $(x_0, y_0)...(x_n, y_n)$ over 2×2, which will be denoted by $x : y$ [6].

Lemma 2.4. *For any transducer $\mathcal{A}$ and a state t, there exists a DFA $\mathcal{A}(t)$ over the alphabet 2×2 such that the language $\mathcal{A}(t)$ accepts is exactly*

$$\{x : y \text{ such that } y = x\underline{t}\}$$

Proof. Consider each transition a/a or $a/\bar{a}$ in $\mathcal{A}$ as a character in 2×2. Add a garbage state g to $\mathcal{A}$, and for each state s add the transitions a/a to g if s is a toggle state, and add the transitions $a/\bar{a}$ to g if s is a copy state. Finally, add all possible transitions from g to itself so that it is a self loop. Thus, the resulting machine $\mathcal{A}(t)$ is a DFA over 2×2 by choosing the start state to be t and the end states to be all states but g.
It can be shown by induction on length of strings over 2×2 that the language recognized by $\mathcal{A}(t)$ is exactly the desired set. □

Definition 2.9. The DFA built as above is referred to as the *acceptor of $\mathcal{A}$ at t* and is denoted as $\mathcal{A}(t)$. Minimality as defined in the previous section is assumed on the acceptor.

Definition 2.10. For two transducers $\mathcal{A}_1$ and $\mathcal{A}_2$, the *product transducer* is the transducer $\mathcal{A}$ with states (s_1, s_2) for each state s_1 in $\mathcal{A}_1$ and s_2 in $\mathcal{A}_2$, and transitions

$$\partial_a\underline{(s_1, s_2)} = \left(\partial_a\underline{s_1}, \partial_{a\underline{s_1}}\underline{s_2}\right)$$

and (s_1, s_2) is a toggle state if and only if exactly one of s_1, s_2 is a toggle state.

Lemma 2.5. *For each state (s_1, s_2) in the product transducer $\mathcal{A}_1 \times \mathcal{A}_2$,*

$$\underline{(s_1, s_2)} = \underline{s_1 s_2}$$

holds true as elements of $Aut(2^*)$.

Proof. Proof by induction on length of binary strings. □

Theorem 2. *The problem of whether $\mathcal{S}(\mathcal{A})$ is commutative or not can be checked in $O(mn^2 \log n)$, where m is the number of toggle states and n is the number of states in $\mathcal{A}$.*

Proof. By Theorem 1, it suffices to check that every copy state transitions into a unique state, and that for all toggle states t, $(\partial_1\underline{t})(\partial_0\underline{t})^{-1} = \theta_{\mathcal{A}}$ is a unique function. The first condition can clearly be checked in $O(n)$. For the second condition, note that for each toggle state t_i,

$$(\partial_0\underline{t_i})^{-1} = \partial_1\underline{t_i^{-1}}$$

in the inverse transducer $\mathcal{A}^{-1}$. So consider the DFA's

$$T_i = (\mathcal{A} \times \mathcal{A}^{-1})\left((\partial_1\underline{t_i}, \partial_1\underline{t_i^{-1}})\right)$$

constructed for each toggle state t_i according to the two lemmas above. The language accepted by T_i is exactly

$$\{x : y \text{ such that } y = x(\partial_1\underline{t_i})(\partial_0\underline{t_i})^{-1}\}$$

so DFA's must be equivalent to each other if there is a unique shift function $\theta_{\mathcal{A}}$. Each T_i has roughly n^2 states, so two DFA's T_i and T_j can be tested for equivalence in $O(n^2 \log(n^2)) = O(n^2 \log n)$ by a variation of Hopcroft's minimization algorithm. Since it is required that all T_i's are equivalent for commutativity of $\mathcal{S}(\mathcal{A})$, it takes $O(mn^2 \log n)$ (under uniform cost function) to test the equivalence sequentially. □

3 Abelian Transduction Groups

This section treats the subgroup $\mathcal{G}(\mathcal{A})$ of $Aut(2^*)$ generated by the associated automorphisms of an invertible binary transducer. Transducers in this section are assumed to be minimal. The main result of this section is an algorithm for determining whether $\mathcal{S}(\mathcal{A}) = \mathcal{G}(\mathcal{A})$. This answers one of the open decidability questions posed in [4], along with the results from the previous section. In order to answer questions regarding the abelian group $\mathcal{G}(\mathcal{A})$, it is natural to consider $\mathbb{Z}$-module homomorphisms represented as integer matrices. Another focus of this section is to expand upon the linear algebraic methods developed by Nekrashevych and Sidki in [3]. The most important results from that paper are outlined below.

Theorem 3. *If $\mathcal{A}$ is a commutative transducer, then $\mathcal{G}(\mathcal{A})$ is isomorphic to either a finite Boolean group, or to $\mathbb{Z}^m$ for some $m \geq 1$. In the latter case, there is an isomorphism $\phi : \mathcal{G}(\mathcal{A}) \to \mathbb{Z}^m$ satisfying the following recursion*

$$\phi^{-1}(v) = \begin{cases} (\phi^{-1}(A \cdot v), \phi^{-1}(A \cdot v)) & \text{if } \phi^{-1}(v) \text{ is even} \\ (\phi^{-1}(A \cdot v - r), \phi^{-1}(A \cdot v + r))\sigma & \text{else} \end{cases}$$

where $A \in GL(m, \mathbb{Q})$ and $v \in \mathbb{Q}^m$ are unique and satisfy for all $v \in \mathbb{Z}^m$ either $A \cdot v \in \mathbb{Z}^m$ or $A \cdot v \pm r \in \mathbb{Z}^m$.

Definition 3.1. The matrix A above is referred to as the *residue matrix* of $\mathcal{A}$ and the vector r above is referred to as the *residue vector*.

The following properties on A are deduced.

Theorem 4. *If $\mathcal{G}(\mathcal{A}) \cong \mathbb{Z}^m$ and A is its associated residual matrix, it satisfies the following*

1. *A is contracting, i.e., its spectral radius is strictly less than 1.*
2. *A is $1/2$-integral, meaning that $A^{-1}(\mathbb{Z}) \cap \mathbb{Z}^m$ is a subgroup of index 2 in $\mathbb{Z}^m$. Therefore A may be represented as*

$$\begin{bmatrix} \frac{a_{1,1}}{2} & a_{1,2} & \cdots & a_{1,m} \\ \frac{a_{2,1}}{2} & a_{2,2} & \cdots & a_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{a_{m,1}}{2} & a_{m,2} & \cdots & a_{,,m} \end{bmatrix}$$

   where all $a_{i,j}$ are integers.

3. *The characteristic polynomial $\chi_A(x)$ is irreducible over $\mathbb{Q}$, and has the form*

$$\chi_A(x) = x^m + \frac{1}{2}g(x)$$

   for some $g \in \mathbb{Z}[x]$ of degree $m-1$. In particular, the constant term is $\pm\frac{1}{2}$.

4. *A is invertible, and the characteristic polynomial $\chi_{A^{-1}}(x)$ is integral and irreducible over $\mathbb{Q}$. From property 2, Laplace expansion yields that A^{-1} is an integral matrix that is similar to the companion matrix of $\chi_{A^{-1}}(x)$ over $\mathbb{Q}$.*

Latimer and MacDuffee proved the following theorem.

Theorem 5. *If $p(x) \in \mathbb{Z}[x]$ is monic and irreducible, the $GL(m, \mathbb{Z})$ similarity classes of integral matrices whose characteristic polynomial coincides with $p(x)$ is in one-to-one correspondance with ideal classes of the ring $\mathbb{Z}[\theta]$, where θ is any root of $p(x)$.*

Property 1 of 4 gives a bound on the coefficients of $\chi_A(x)$, and combined with property 4 and Theorem 5, it can be shown that there are only finitely many possibilities of A per m, upto $GL(m, \mathbb{Z})$ similarity.

Definition 3.2. A residue matrix A is *well-behaved* if it has only one $GL(m, \mathbb{Z})$ similarity class.

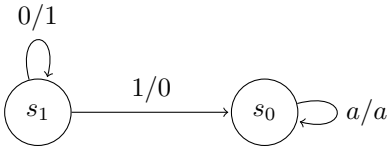

Figure 2: The adding machine

Nekrashevych and Sidki's computation implies that most residue matrices are well-behaved.
Now, consider the following example,

Example 3.1. *There exist commutative transducers for which $\mathcal{S}(\mathcal{A})$ is not a group. The adding machine is such an example, as $\mathcal{S}(\mathcal{A}) \cong \mathbb{N}$, but $\mathcal{G}(\mathcal{A}) \cong \mathbb{Z}$.*

This shows that the semigroup $\mathcal{S}(\mathcal{A})$ may not be closed under inverses. However, a polynomial time algorithm for computing the rank of $\mathcal{G}(\mathcal{A})$ and (A, r) pair will be presented. With this information, it is easy to construct a transducer whose transduction semigroup coincides with $\mathcal{G}(\mathcal{A})$.

Theorem 6. *$\mathcal{G}(\mathcal{A})$ is a finite Boolean group iff every state transitions into a unique state.*

Proof. Right to left is immediate from the previous results.
For the left to right direction, it suffices to show that the shift function of $\mathcal{A}$ is the identity. Let t be a toggle state, so that

$$I = \underline{t}^2$$

Evaluating ∂_0 on both sides yields,

$$\begin{aligned} I &= (\partial_0 \underline{t})(\partial_1 \underline{t}) \\ &= (\partial_0 \underline{t})^2 \theta_{\mathcal{A}} \\ &= \theta_{\mathcal{A}} \end{aligned}$$

□

Corollary 3.1. *There is an $O(n)$ algorithm to determine if $\mathcal{G}(\mathcal{A})$ is a finite abelian group, where n is the number of states in $\mathcal{A}$. In other words, the strongly connected components are disjoint cycles.*

Proof. Check that every state transitions to a unique state. □

Now consider the case when $\mathcal{G}(\mathcal{A})$ is a free abelian group. The case when $\mathcal{G}(\mathcal{A}) \cong \mathbb{Z}$ is completely characterized by Nerkrashevych and Sidki. Hence, assume that $\mathcal{G}(\mathcal{A})$ is torsion free of rank at least 2, say $\mathcal{G}(\mathcal{A}) \cong \mathbb{Z}^m$ via an isomorphism

$$\phi : \mathcal{G}(\mathcal{A}) \to \mathbb{Z}^m$$

Let $\mathcal{A}$ have n states $s_1, ..., s_n$, where the first t states are the toggle states. Then there is a natural homomorphism $\psi : \mathbb{Z}^n \to \mathcal{G}(\mathcal{A})$ where

$$\psi(a_1, ..., a_n) = \underline{s_1}^{a_1} ... \underline{s_n}^{a_n}$$

Let H be the subgroup of $\mathcal{G}(\mathcal{A})$ consisting of even functions, and let $V = \phi(H)$ in $\mathbb{Z}^m$. Then V must have index 2, so $V \cong (2\mathbb{Z}) \oplus \mathbb{Z}^{m-1}$. Then the linear map A associated to $\mathcal{A}$ acts as a homomorphism from V to $\mathbb{Z}^m$ mimicking the behavior of ∂_0, so the following diagram commutes.

1

$$\begin{array}{ccc} \mathcal{G}(\mathcal{A}) & \xrightarrow{\phi} & \mathbb{Z}^m \\ \uparrow{\scriptstyle \partial_0} & & \uparrow{\scriptstyle A} \\ H & \xrightarrow{\phi} & V \end{array}$$

Now, let

$$U = (2\mathbb{Z})^t \oplus \mathbb{Z}^{n-t}$$

be a submodule of $\mathbb{Z}^n$. Note that $\psi(U) \leq H$, and consider the $n \times n$ matix B defined by

$$B_{i,j} = \begin{cases} 1 & \text{if } \partial_0 \underline{s_i} = \partial_1 \underline{s_i} = \underline{s_j} \\ \frac{1}{2} & \text{if } s_j \text{ is a toggle state, and } \partial_0 \underline{s_j} = \underline{s_i}, \text{ or } \partial_1 \underline{s_j} = \underline{s_i} \\ 0 & \text{otherwise} \end{cases}$$

which is the transpose of the transition matrix of the system. Note that the matrix is well-defined by the assumption that $\mathcal{G}(\mathcal{A})$ is torsion-free (hence no toggle state transitions into a unique state), and that $B(U) \subseteq \mathbb{Z}^n$.

Proposition 3.1.

$$\psi \circ B = \partial_0 \circ \psi$$

as functions from U to $\mathcal{G}(\mathcal{A})$.

Proof. Obvious from the exponentiation rule shown in the previous section. □

As a consequence, the following diagram commutes :

$$\begin{array}{ccccc} \mathbb{Z}^n & \xrightarrow{\psi} & \mathcal{G}(\mathcal{A}) & \xrightarrow{\phi} & \mathbb{Z}^m \\ \uparrow{\scriptstyle B} & & \uparrow{\scriptstyle \partial_0} & & \uparrow{\scriptstyle A} \\ U & \xrightarrow{\psi} & H & \xrightarrow{\phi} & V \end{array}$$

Consider the linear map $R = \phi \circ \psi$ which has a representation as a $m \times n$ integral matrix.

Proposition 3.2. *R has full rank.*

Proof. ϕ is a bijection and ψ is a surjection. □

Proposition 3.3.

$$RB = AR$$

as functions from U to $\mathbb{Z}^m$.

Proof. Obvious from the commutative diagram. □

This gives an important relationship of eigenvalues of A and B.

Lemma 3.1. *The characteristic polynomial of A divides the characteristic polynomial of B.*

Proof. From the relation $RB = AR$, it is easy to show by induction that for any $k \geq 1$,

$$RB^k = A^k R$$

and it follows that for any polynomial P,

$$RP(B) = P(A)R$$

Hence, if χ_B is the characteristic polynomial of B, Cayley-Hamilton theorem gives

$$\begin{aligned} 0 &= \chi_B(B) \\ &= R\chi_B(B) \\ &= \chi_B(A)R \end{aligned}$$

and since the rank of R is m, it follows that the rank of $\chi_B(A)$ is 0 and hence $\chi_B(A) = 0$. Since the characteristic polynomial χ_A of A is irreducible, it is also the minimal polynomial of A, and hence χ_A divides χ_B. □

Theorem 7. *There is a polynomial-time algorithm that computes the correct matrix A, the residuation vector r and a basis of identities of $\mathcal{G}(\mathcal{A})$.*

Proof. For each irreducible factor f of χ_B of the form

$$f(x) = x^m + \frac{1}{2}g(x)$$

where $g(x) \in \mathbb{Z}[x]$, let C_f be its companion matrix. Solve the system of equations

$$TRB = C_f TR$$

where R must be integral and T must be an invertible matrix over $\mathbb{Q}$ such that

$$T^{-1}C_f T$$

is 1/2-integral (so T may also be specified to be integral). If it is not solvable, try a different irreducible factor. Since A must be $GL(m, \mathbb{Q})$ similar to C_f if f were chosen correctly, the algorithm has at least one solution and is unique upto $GL(m, \mathbb{Z})$ similarity.
When the matrices A and R are known, r can be computed from the automaton $\mathcal{A}$ by looking at the transition of an odd state.
Factoring $\chi_B(x)$ can be done by applying the LLL algorithm which runs in roughly $O(n^7)$ time. There are $O(n)$ many possible candidates for $\chi_A(x)$ given the factorization, and for each candidate, computing the integer matrix R that satisfies $RB = AR$ (or lack thereof) takes $O(n^3)$ time. □

This gives the following decision algorithm. Note that this problem is trivially semidecidable but decidability is far from obvious.

Theorem 8. *If $\mathcal{S}(\mathcal{A})$ is commutative, the problem of $\mathcal{S}(\mathcal{A})$ being a group is decidable.*

Proof. Let K denote the matrix whose columns are the integer basis vectors of $\ker(R)$ as in the previous algorithm. It is easy to see that

$$K \cdot e > 0$$

has a rational solution iff $\mathcal{S}(\mathcal{A}) = \mathcal{G}(\mathcal{A})$. □

4 Canonical automaton

The canonical automaton is a transducer defined per transition matrix. It will be shown that any transducer can be written as an extension of some canonical automaton. A canonical automaton can also be interpreted as the intersection of all transducers defined with some fixed transition matrix. A conjecture related to the structure of the canonical automaton is also presented, along with partial results and outline of attempts.

Definition 4.1. Let A be a residue matrix, so that $e_1, ..., e_m$ are the basis vectors cycled by A ($A \cdot e_k = e_{k+1}$ for all valid k). Then the *canonical automaton* for A is the automaton associated to the pair (A, r) where $r = A \cdot e_1$, generated at e_1.

An important observation is that the choice of r and starting at e_1 immediately gives 0 as a state since $A \cdot e_1 - r = 0$. This results in a sink strongly connected component at 0. Computation implies that the residues of e_1 starting at $A \cdot e_1 + r = 2r$ forms a SCC by itself as well. This question will be explored in the next section.
Recall that a residue matrix is well-behaved if it has only one $GL(m, \mathbb{Z})$-similarity class and therefore it can assumed to be equal to the companion matrix of its characteristic polynomial. For the rest of this section, A will be assumed to be well-behaved and r will always refer to $A \cdot e_1$, where $e_1, ..., e_m$ refer to the canonical basis vectors. $\mathcal{A}$ will refer to the canonical automaton whenever A is specified. Any residuation vector, in general, will be denoted by r'.

Definition 4.2. Let (A, r') be a residuation pair. Then the *complete automaton* is the infinite automaton with $\mathbb{Z}^m$ as its state set and the transitions defined by (A, r') [3].

It is necessary to introduce basic notions related to analysis of vectors and matrices. The following propositions specify the entries of A and its related matrices.

Proposition 4.1. *Consider the companion matrix of a monic polyomial* $P(x) = x^m + c_{m-1}x^{m-1} + ... + c_1 x + c_0$ *which is*

$$C = \begin{bmatrix} -c_{m-1} & 1 & & \\ \vdots & & \ddots & \\ -c_1 & & & 1 \\ -c_0 & 0 & \dots & 0 \end{bmatrix}$$

then its inverse is given by

$$C^{-1} = \begin{bmatrix} 0 & \dots & 0 & -\frac{1}{c_0} \\ 1 & & & -\frac{c_{m-1}}{c_0} \\ & \ddots & & \vdots \\ & & 1 & -\frac{c_1}{c_0} \end{bmatrix}$$

which satisfies $C^{-1} = PC^*P^{-1}$ *where* P *is the permutation matrix*

$$P = P^{-1} = \begin{bmatrix} & & 1 \\ & \iddots & \\ 1 & & \end{bmatrix}$$

and C^* *is the companion matrix to the reciprocal polynomial* $P^*(x) = x^m + \frac{c_1}{c_0}x^{m-1} + ... + \frac{c_{m-1}}{c_0}x + \frac{1}{c_0}$.

Proposition 4.2. *Let* C *be a companion matrix to* $P(x)$ *as defined as above. Let* $\lambda_1, ..., \lambda_m$ *denote the roots of* $P(x)$. *Then for* $1 \le i \le m$, *the vector*

$$[\lambda_i^{m-1} \dots \lambda_i \ 1]^T$$

is a left eigenvector of C *to eigenvalue* λ_i. *Hence, the Vandermonde matrix*

$$V = \begin{bmatrix} \lambda_1^{m-1} & \dots & \lambda_1 & 1 \\ \vdots & \ddots & \vdots & \\ \lambda_m^{m-1} & \dots & \lambda_m & 1 \end{bmatrix}$$

satisfies

$$VAV^{-1} = \Lambda$$

where Λ *is the diagonal eigenvalue matrix.*

It is also useful to agree on a norm that captures the eigenvalue information of the matrix A.

Proposition 4.3. *Let* λ *denote the spectral radius of* A*. Since* V *is invertible, the function* $\|\cdot\| : \mathbb{C}^m \to \mathbb{R}$ *defined by* $\|x\| = \|V \cdot x\|_\infty$ *is a norm (*$\|\cdot\|_\infty$ *is the* ∞*-norm). This norm satisfies* $\|r\| = \lambda^m$ *and the induced matrix norm for* $\|\cdot\|$ *satisfies* $\|A\| = \lambda$*.*

Proof.

$$\begin{aligned} \|r\| &= \|Vr\|_\infty \\ &= \|VAe_1\|_\infty \\ &= \|\Lambda V e_1\|_\infty \\ &= \max_i\{\lambda_1^m, ..., \lambda_m^m\} \\ &= \lambda^m \end{aligned}$$

and

$$\begin{aligned} \|A\| &= \sup\{\|A \cdot x\| \mid \|x\| = 1\} \\ &= \sup\{\|\Lambda V \cdot x\|_\infty \mid \|V \cdot x\|_\infty = 1\} \\ &= \sup\{\|\Lambda \cdot y\|_\infty \mid \|y\|_\infty = 1\} \\ &= \|\Lambda\|_\infty \\ &= \lambda \end{aligned}$$

□

Let (A, r') be any residue pair, and note that r' can be written as $r + u$ for some $u \in \mathbb{Z}^m$. It is easy to see that the canonical automaton $\mathcal{A}$ is embedded in the complete automaton associated with the pair (A, r') by taking $v = A^{-1} \cdot r'$ to be the generating vector. The following theorem identifies this monomorphism.

Theorem 9. *Let* (A, r') *be any residuation pair and let* $r = A \cdot e_1$*. Consider following matrix*

$$\rho = \begin{bmatrix} A^{-1}r' & \dots & A^{-m}r' \end{bmatrix}$$

which is integral. Then ρ *satisfies the following properties :*

1. $A \cdot \rho(v) = \rho(A \cdot v)$
2. $A \cdot \rho(v) \pm r' = \rho(A \cdot v \pm r)$

Proof. It suffices to show that ρ and A commute, since then

$$\begin{aligned} \rho(r) &= \rho(A \cdot e_1) \\ &= A\rho(e_1) \\ &= AA^{-1}r' \\ &= r' \end{aligned}$$

and the two properties follow by linearity.

Note that

$$A^{-1} \cdot r' = A^{-1}(r + u) = e_1 + A^{-1}u$$

hence

$$\rho = \begin{bmatrix} e_1 + A^{-1}u & \dots & e_m + A^{-m}u \end{bmatrix} = I_m + \begin{bmatrix} A^{-1}u & \dots & A^{-m}u \end{bmatrix}$$

Let $F : \mathbb{Z}^m \to Mat_{m\times m}(\mathbb{Z})$ be defined by

$$F(u) = \begin{bmatrix} A^{-1}u & \dots & A^{-m}u \end{bmatrix}$$

To show that ρ commutes with A, it suffices to show that A commutes with $F(u)$. Since F is a $\mathbb{Z}$-module homomorphism, it suffices to show that for each basis vector e_k, A commutes with $F(e_k)$. Finally,

$$\begin{aligned} F(e_k) &= \begin{bmatrix} A^{-1}e_k & \dots & A^{-m}e_k \end{bmatrix} \\ &= A^{-k}\begin{bmatrix} A^{-k-1}e_k & \dots & A^{-km}e_k \end{bmatrix} \\ &= A^{-k} \end{aligned}$$

which obviously commutes with A. □

5 Single Sink SCC Conjecture

The following conjecture results from inspecting the topology of canonical automata for distinct residue matrices A.

Conjecture 5.1 (Single Sink SCC Conjecture)**.** *The canonical automaton has a single sink strongly connected component consisting of a self loop at* 0*.*

This section describes some attempts at proving Conjecture **??**. The following is another conjecture that would imply Conjecture 5.1.

Conjecture 5.2. *For any automaton $\mathcal{A}'$, it has at most 1 sink strongly connected component.*

There are mainly two approaches taken at proving the conjecture. The first attempt is analytic, in which residuation is expressed via a matrix polynomial and the $\|\cdot\|$ norm defined in the previous section is used to derive bounds. First, some new definitions must be introduced.

Definition 5.1. For any $u, v \in \mathbb{Z}^m$, if there is some word $x \in 2^*$ such that $\partial_x v = u$, then v is said to be an *ancestor* of u and u is said to be a *descendant* of v. In other words, u and v are connected by a directed path.

Residuation in a transducer can be expressed via a matrix equation relating an ancestor to its descendant.

Proposition 5.1. *For $u, v \in \phi(\mathcal{G}(\mathcal{A}))$, u is an ancestor of v if and only if there exists some $n \geq 0$ and some $P(x)$ with coefficients from $\{+1, 0, -1\}$ of degree strictly less than n such that*

$$u = A^{-n} \cdot v + P(A^{-1}) \cdot v_1$$

Proof. Both directions are done by induction on the degree n. □

The set of integer polynomials with coefficients from $\{+1, 0, -1\}$ has also been studied by Bloch and Polya in [9].

Proposition 5.2. *For $u, v \in \phi(\mathcal{G}(\mathcal{A}))$, if u is a descendant of v then there exists some $n \geq 1$ and some $P(x) \in \{-1, 0, +1\}[x]$ of degree strictly less than n such that*

$$u = A^n \cdot v + AP(A) \cdot v_1$$

Proof. This is again verified by induction on the degree of descendancy.

□

Note that the second proposition is not an if and only if statement and the subtle differences in the requirements for n. The above observations give the proof for the following lemmae.

Lemma 5.1. *Any complete automaton associated to (A, r') has only finitely many strongly connected components, each of which has finitely many states.*

Proof. Let $v \in \mathbb{Z}^m$. Note that all of its descendants may be written as

$$A^n v + \sum_{k=0}^{n-1} c_k A^k r'$$

where $c_k \in \{-1, 0, 1\}$. With $\|\cdot\|$ denoting the $\|\cdot\|_\infty$ norm defined on $\mathbb{Q}^m$ and the induced matrix norm interchangeably, triangle inequality gives the bound

$$\|A^n\|\|v\| + \left(\sum_{k=0}^{n-1} \|A^k\|\right) \|r'\|$$

on the norm of descendants of v. Recall that $\|A^k\| = \lambda^k$ where λ is the specrtral radius of A. Therefore, taking $n \to \infty$ in the above expression gives

$$\delta = \left(\sum_{k=0}^{\infty} \|A^k\|\right) \|r'\|$$

which is independent of $\|v\|$. This has the implication that eventually, all descendants of v must have norm bounded by δ. Since the ball of radius δ around 0 in $\mathbb{Z}^m$ with respect to $\|\cdot\|$ norm is finite, this implies that there can only be finitely many strongly connected components, each of which has finitely many states. □

Lemma 5.2. *Consider a state s in the canonical automaton, and suppose that there exists some $x \in 2^*$ such that $\partial_x \underline{s} = \underline{s}$. If $u = \phi(\underline{s})$, then $\|u\| \leq \frac{\lambda^m}{1-\lambda}$ must be true, where $\|\cdot\|$ is the norm defined earlier in this section.*

Proof. Suppose that x is of length n. Then for all $k \geq 1$, there exists some $P_k \in \{-1, 0, 1\}[x]$ of degree at most $kn - 1$ such that

$$u = A^{kn} u + P_k(A) r$$

Recall that the induced matrix norm $\|\cdot\|$ is sub-multiplicative and compatible, and that

$$\|A\| = \lambda, \|r\| = \lambda^m$$

where λ is the spectral radius of A. Thus,

$$\begin{aligned} \|u\| &\leq \|A\|^{kn}\|u\| + \left(\sum_{i=0}^{kn-1} \|A\|^i\right) \|r\| \\ &\leq \lambda^{kn}\|u\| + \left(\frac{1}{1-\lambda}\right) \lambda^m \end{aligned}$$

by taking $\lim k \to \infty$, this gives

$$\|u\| \leq \frac{\lambda^m}{1-\lambda}$$

□

However, extending the above approach to prove statements relating to Conjecture 5.1 has turned out to be difficult. The second approach is more automata theoretic and considers the behavior of individual vectors as automorphisms, rather than vectors.

Definition 5.2. A function $f \in \mathcal{G}(\mathcal{A})$ is *tame* if there is some word $x \in 2^*$ such that $\partial_x f = Id$.

Definition 5.3. A function $f \in \mathcal{G}(\mathcal{A})$ is *completely tame* if all its descendants are tame.

Proposition 5.3. *The product of a tame function and a completely tame function is tame.*

Proof. Let f be tame, and g be completely tame. Then there exists some word $x \in 2^*$ such that $\partial_x f = Id$ and since g is completely tame, there is some word y such that $\partial_y(\partial_{xf} g) = Id$. Now, consider $z = x : y$. Then

$$\partial_z(fg) = (\partial_z f)(\partial_{zf} g)$$

Note that $\partial_z f = \partial_y(\partial xf) = Id$ and $zf = (xf) : (y\partial_x f) = (xf) : y$. Then $\partial_{zf} g = \partial_y(\partial_{xf} g) = Id$, so this gives $\partial_z(fg) = Id$. □

Definition 5.4. A subgroup of $\mathcal{G}(\mathcal{A})$ is said to be state-closed if for all $f \in \mathcal{G}(\mathcal{A})$, $\partial_a f \in \mathcal{G}(\mathcal{A})$ for $a \in \{0, 1\}$.

Lemma 5.3. *Completely tame functions of $\mathcal{G}(\mathcal{A})$ forms a state-closed subgroup of $\mathcal{G}(\mathcal{A})$.*

Proof. It is easy to see that the set of completely tame functions of $\mathcal{G}(\mathcal{A})$ contains the identity and is closed under inverse and residues. Now, let f, g be two completely tame functions, and let $x \in 2^*$. Then

$$\partial_x(fg) = (\partial_x f)(\partial_{xf} g)$$

where $\partial_x f$ and $\partial_{xf} g$ are both completely tame. By the previous proposition, $\partial_x(fg)$ is shown to be tame. □

Let $f \in \mathcal{G}(\mathcal{A})$ be a non-identity function. Then the residuation pair (A, r) still computes the descendants of f, which are also elemends of $\mathcal{G}(\mathcal{A})$. Hence it is natural to consider the automaton $\mathcal{A}_f$ generated by f under residuation, which is finite since A is a contraction. The transduction semigroup $\mathcal{S}(\mathcal{A}_f)$ is a subsemigroup of $\mathcal{S}(\mathcal{A})$ and hence $\mathcal{G}(\mathcal{A}_f) \leq \mathcal{G}(\mathcal{A})$. Nekrashevych-Sidki proved the following :

Proposition 5.4. *If $f \in \mathcal{G}(\mathcal{A})$ is not the identity function, then the image of $\mathcal{G}(\mathcal{A}_f)$ under the isomorphism $\phi : \mathcal{G}(\mathcal{A}) \to \mathbb{Z}^m$ is a sublattice of rank m.*

Now, since A is $GL(m, \mathbb{Q})$ similar to the companion matrix for $\chi_A(x)$, there exists some vector $v \in \mathbb{Q}^m$ such that $\{v, A \cdot v, ..., A^{m-1} \cdot v\}$ is a basis for the vector space. By multiplying by some large constant, these vectors can be assumed to lie in $\mathbb{Z}^m$. Further, if there exist rational coefficients $c_1, ..., c_m$ such that

$$\sum_{k=1}^{m} c_k A^k \cdot v = 0$$

then multplying by A^{-1} shows that all $c_k = 0$, so it can further be assumed that $A^m \cdot v \notin \mathbb{Z}^m$. Hence, call $v_1 = A^{m-1} \cdot v, v_2 = A^{m-2} \cdot v, ..., v_m = v$ and set $r = A \cdot v_1$. For the remainder of this section the *canonical* automaton $\mathcal{A}$ generatred from v_1 via the transition pair (A, r) will be considered. The following result is somewhat related to the previous theorem.

Theorem 10. *For f induced by some state of $\mathcal{A}$ and not the identity function, $\mathcal{G}(\mathcal{A}_f) = \mathcal{G}(\mathcal{A})$.*

Proof. As f is not the identity function, taking its residues eventually yields an odd function g. Note that $g \in \mathcal{G}(\mathcal{A}_f)$ and $h = \partial_0 g \in \mathcal{G}(\mathcal{A}_f)$ as well. Let $v = \phi(g)$ and $u = \phi(h)$. Since A^{-1} is integral,

$$v - A^{-1} \cdot u = v - A^{-1} \cdot (A \cdot v - r) = v - A^{-1} \cdot (A \cdot v - A \cdot v_1) = v_1$$

is in $\phi(\mathcal{G}(\mathcal{A}_f))$. So $\mathcal{G}(\mathcal{A}_f)$ contains all descendants of v_1 as elements, implying that $\mathcal{G}(\mathcal{A}_f) = \mathcal{G}(\mathcal{A})$. □

Proposition 5.4 implies that Conjecture 5.1 is equivalent to proving that a nontrivial completely tame function exists. In any case, finding a proof for the conjecture remains to be a significant task.

6 Knuth normal form

In this section transducers are assumed to be commutative and that the transduction groups are torsion free. Further, $\mathcal{S}(\mathcal{A}) = \mathcal{G}(\mathcal{A})$ can be assumed since the automaton $\mathcal{A}'$ constructed from A and r has the property that $\mathcal{S}(\mathcal{A}') = \mathcal{G}(\mathcal{A}')$, and $\mathcal{G}(\mathcal{A}') = \mathcal{G}(\mathcal{A})$.
Recall that a commutative transducer $\mathcal{A}$ with t toggle states can be characterized by the lengths of the copy chains attached to each toggle state, and the two transitions from each toggle state. First, it is useful to define the following functions and conventions.

Definition 6.1. The states of $\mathcal{A}$ will be labeled as follows:

$$s(i,0) \text{ is the } i\text{th toggle state}$$

$$s(i,j) \text{ is the copy state transitioning into } s(i,j-1)$$

Definition 6.2. For $1 \leq i \leq t$ and $a \in \{0,1\}$, let

$$\tau_a(i) = j \text{ and } \delta_a(i) = k \text{ if and only if } \partial_a \underline{s(i,0)} = \underline{s(j,k)}$$

Definition 6.3. For $1 \leq i \leq t$, let

$$l(i) = \max\{j : s(i,j) \in \mathcal{A}\}$$

$$h(i) = \max\{j : \exists 1 \leq k \leq t, a \in \{0,1\} \text{such that } \partial_a \underline{s(k,0)} = \underline{s(i,j)}\}$$

Note that $h(i) \leq l(i)$ always. Finally,

Definition 6.4. For $n \in \mathbb{N}$, define a transducer $\mathcal{A}_n$ to have all the same toggle states and transitions as $\mathcal{A}$, and for each $1 \leq i \leq t$, the length of the copy chain is $h(i) + n$. Further, let $\mathcal{G}_n = \mathcal{G}(\mathcal{A}_n)$.

With the following lemma, it turns out that $\mathcal{A} = \mathcal{A}_0$ can be assumed in general.

Lemma 6.1. *If $\mathcal{G}_n = \mathcal{G}(\mathcal{A})$ for all $n \in \mathbb{N}$.*

Proof. For simplicity assume that $\mathcal{G}_0 \cong \mathbb{Z}^m$, and let A be the $m \times m$ matrix which acts as the ∂_0-function on $(2\mathbb{Z}) \otimes \mathbb{Z}^{m-1}$. It is known that A is $GL(m,\mathbb{Z})$-similar to the companion matrix of its characteristic polynomial, and that the characteristic polynomial of $\mathcal{A}^{-1}$ is integral. It follows that A^{-1} is also $GL(m,\mathbb{Z})$ similar to the companion matrix of its characteristic polynomial, so A^{-1} is an integral matrix. Let v be the image of the automorphism defined at any state s in $\mathcal{A}_0$. Then $A^{-1} \cdot v \in \mathbb{Z}^m$, so the automorphism defined at a copy state transitioning into s belongs to $\mathcal{G}_0$. This can be repeated for any finite number of steps, so all $\mathcal{G}_n$ are equivalent to $\mathcal{G}_0$.
Finally, as there must exist n and n' such that $G_n \leq \mathcal{G}(\mathcal{A}) \leq \mathcal{G}_{n'}$, it follows that all $\mathcal{G}_n$ are equivalent to $\mathcal{G}(\mathcal{A})$. □

Now, consider the transducer $\mathcal{A}_\omega$ obtained by attaching an infinite tail of copy states to each copy chain in $\mathcal{A}$. By previous lemma, $\mathcal{G}(\mathcal{A}_\omega) = \mathcal{G}(\mathcal{A})$, but the following lemma is obtained.

Lemma 6.2. *Let $1 \leq i \leq t$ and $j \geq 0$, then*

$$\underline{s(i,j)}^2 = \underline{s(\tau_0(i), \delta_0(i) + j + 1)}\ \underline{s(\tau_1(i), \delta_1(i) + j + 1)}$$

Proof. By induction on j. For simplicity, denote $\tau_a = \tau_a(i), \delta_a = \delta_a(i)$.
Base case :

$$\underline{s(i,0)}^2 = (\underline{s(\tau_0, \delta_0)}\ \underline{s(\tau_1, \delta_1)}, \underline{s(\tau_0, \delta_0)}\ \underline{s(\tau_1, \delta_1)})$$

and

$$\partial_0 \underline{s(\tau_0, \delta_0 + 1)} = \partial_1 \underline{s(\tau_0, \delta_0 + 1)} = \underline{s(\tau_0, \delta_0)}$$

and

$$\partial_0 \underline{s(\tau_1, \delta_1 + 1)} = \partial_1 \underline{s(\tau_1, \delta_1 + 1)} = \underline{s(\tau_1, \delta_1)}$$

proves the base case.
Induction step : Assume that the claim is true for some $j \geq 0$. Then

$$\underline{s(i,j+1)}^2 = (\underline{s(i,j)}^2, \underline{s(i,j)}^2)$$

where by induction step,

$$\underline{s(i,j)}^2 = \underline{s(\tau_0, \delta_0 + j + 1)}\ \underline{s(\tau_1, \delta_1 + j + 1)}$$

Now,

$$\partial_0 \underline{s(\tau_0, \delta_0 + j + 2)} = \partial_1 \underline{s(\tau_0, \delta_0 + j + 2)} = \underline{s(\tau_0, \delta_0, j + 1)}$$

and

$$\partial_0 \underline{s(\tau_1, \delta_1 + j + 2)} = \partial_1 \underline{s(\tau_1, \delta_1 + j + 2)} = \underline{s(\tau_1, \delta_1, j + 1)}$$

proves this case. □

The previous lemma can be interpreted as a rewrite rule for elements of $\mathcal{G}(\mathcal{A}_\omega)$. If it terminates for a representation of a function f in $\mathcal{G}(\mathcal{A})$, then it must look like

$$f = \underline{s(i_1, j_1)}...\underline{s(i_r, j_r)}$$

where (i_k, j_k) are unique.

Lemma 6.3. *The above rewrite process terminates.*

Proof. For a representation

$$f = f_1^{k_1}...f_r^{k_r}$$

of an automophism $f \in \mathcal{G}(\mathcal{A})$, where each f_i is the automorphism defined at some state, let

$$\sum e_i$$

be the weight of that expression. □

It is helpful to define the following directed graph

Definition 6.5. For a commutative transducer $\mathcal{A}$, define the *toggle graph* of $\mathcal{A}$, $Gr(\mathcal{A})$, with t vertices $v_1, ..., v_t$ and a directed edge (v_i, v_j) exists iff either

$$\partial_0 \underline{s(i,0)} = \underline{s(j,k)}$$

or

$$\partial_1 \underline{s(i,0)} = \underline{s(j,k)}$$

for some k.

Proposition 6.1. *A rewrite process for an expression of a function in $\mathcal{G}(\mathcal{A}_\omega)$ does not terminate if and only if there exists an equivalent expression*

$$f = \underline{s(i_1, j_1)}...\underline{s(i_r, j_r)}$$

of the same function such that there exists a strongly connected component of $Gr(\mathcal{A})$ containing all v_{i_k}, and

7 Orbit equivalence

In this section the problem of orbit equivalence is discussed for functions in $\mathcal{G}(\mathcal{A})$. First, some definitions.

Definition 7.1. For $f \in \mathcal{S}(\mathcal{A})$, define the *orbit relation* of f to be the relation

$$f^* = \{(u, v) : u = vf^i \text{ for some } i \geq 0\}$$

on 2^*.

Definition 7.2. For $f \in \mathcal{S}(\mathcal{A})$ and $u \in 2^*$, the *orbit of u under f* is

$$\text{orb}_f(u) = \{uf^i : i \geq 0\}$$

It is easy to see that as elements of $\mathcal{S}(\mathcal{A})$ are length preserving, each orbit is finite. It has also been shown that every orbit has size 2^n for some $n \geq 0$.

Definition 7.3. Two automorphisms f and g are said to be *orbit equivalent* iff

$$\text{orb}_f(u) = \text{orb}_g(u)$$

for all $u \in 2^*$.

Definition 7.4. For $f \in \mathcal{G}(\mathcal{A})$, the *orbit tree* of f is defined to be the quotient $2^*/f^*$.

Proposition 7.1. *Let f and g be functions from $\mathcal{G}(\mathcal{A})$, and let $\mathfrak{T}$ denote the orbit tree of f. Then f and g are orbit equivalent iff for each $x \in \mathfrak{T}$, there is an odd integer c_x such that*

$$xf = xg^{c_x}$$

and this choice of c_x is unique modulo $|orb_f(x)|$.

Proof. Trivial proof. □

Thus, c_x will refer to the least positive integer that satisfies the equation and the following result holds.

Proposition 7.2. *If $orb_f(x) = orb_f(y)$, then $c_x = c_y$.*

Proof. Trivial proof. □

Proposition 7.3. *For any $x \in \mathfrak{T}$,*

$$c_{x0} = c_{x1}$$

holds, and is equal to either c_x or $c_x + |orb_f(x)|$. The latter case occurs iff orbit of x doubles under f and

Proof. Let $|\text{orb}_f(x)| = 2^n$. Then

$$\begin{aligned} x0f &= (xf) \cdot (0\partial_x f) \\ x0g^{c_{x0}} &= (xg^{c_{x0}}) \cdot (0\partial_x g^{c_{x0}}) \end{aligned}$$

implies

$$c_{x0} \equiv c_x \bmod 2^n$$

If orbit of x splits, then $|\text{orb}_f(x0)| = 2^n$ as well, so c_{x0} must be equal to c_x and similarly for c_{x1}. If orbit of x doubles, then

$$\text{orb}_f(x0) = \text{orb}_f(x1)$$

so

$$c_{x0} = c_{x1}$$

□

8 Orbit rationality

The following definition and lemma can be found in [4].

Definition 8.1. Given $f, g \in \mathcal{G}(\mathcal{A})$, the *orbit relation of f with translation g* is defined to be

$$R(f,g) = \{(u,v) : u = vf^i g \text{ for some } i \geq 0\}$$

Lemma 8.1. *(Quotient Lemma) Let $f, h \in \mathcal{G}(\mathcal{A})$ and let $b = ah$ for $a \in \{0,1\}$. If f is even,*

$$(a:b)^{-1}R(f,h) = R(\partial_a f, \partial_a h)$$

If f is odd,

$$(a:b)^{-1}R(f,h) = R(\partial_a f \partial_{\bar{a}} f, \partial_a h)$$

and

$$(a:\bar{b})^{-1}R(f,h) = R(\partial_a f \partial_{\bar{a}} f, \partial_a f \partial_{\bar{a}} h)$$

All other quotients are empty.

Lemma 8.2. *A function $f \in \mathcal{G}(\mathcal{A})$ is orbit rational if and only if the sequence of functions $\{f_k\}_{k\in\mathbb{N}}$ defined by*

$$f_0 = f$$
$$f_{k+1} = \begin{cases} \partial f_k & \text{if } f_k \text{ is even} \\ \partial(f_k^2) & \text{if } f_k \text{ is odd} \end{cases}$$

is finite modulo orbit equivalence.

Proof. By Quotient Lemma, look at the first coordinate. □

By setting $\phi(f) = v \in \mathbb{Z}^m$, the sequence of functions above may be written as sequence of vectors of the form

$$v_t = 2^{s(t)} A^t v$$

where $s(t)$ is the least non-negative integer such that $v_t \in \mathbb{Z}^m$. The following is a special case of the matrix A.

Lemma 8.3. *If there is a power of A with a rational eigenvalue, say A^k has an eigenvalue $r \in \mathbb{Q}$, then $r = \pm\frac{1}{2^{k/m}}$ and*

$$A^k = rI$$

where m divides k.

Proof. Let λ be an eigenvalue of A such that $r = \lambda^k$. Then λ satisfies the rational polynomial

$$p(x) = x^k - r$$

Since $\chi_A(x)$ is irreducible, it is also the minimal polynomial of λ over $\mathbb{Q}$, so $\chi_A(x)$ divides $p(x)$. This implies that all eigenvalues of A satisfy $p(x)$, and it follows that

$$A^k = rI$$

Now this implies that all eigenvalues of A have the same norm. Since the constant term of $\chi_A(x)$ is $\pm\frac{1}{2}$, the norm of any eiganvalue of A must be $2^{-\frac{1}{m}}$. Hence

$$r = \lambda^k = \pm\frac{1}{2^{k/m}}$$

holds true and in particular, m divides k. □

Lemma 8.4. *If there is a power of A such that all its eigenvalues are rational, then every function of $\mathcal{G}(\mathcal{A})$ is orbit rational.*

Proof. By Lemma 8.3, the sequence of vectors

$$\{2^{s(t)}A^t v\}_{t\in\mathbb{N}}$$

is finite for any $v \in \mathbb{Z}^m$ since if $A^k = \frac{1}{2^{k/m}}I$, then $s(k) = k/m$. □

Definition 8.2. If f is even, let $\partial_0 f = \partial_1 f$ be the *residual* of f and denote it by ∂f.

Definition 8.3. For any $f \in \mathcal{G}(\mathcal{A})$, define

$$\nu_2(f) = \begin{cases} \infty & \text{if } f = Id \\ 0 & \text{if } f \text{ is odd} \\ 1 + \nu_2(\partial f) & \text{if } f \text{ is even and not } Id \end{cases}$$

Definition 8.4. For $f \in \mathcal{G}(\mathcal{A})$ and $0 \le k \le \nu_2(f)$, let the *k-th residual of f* be

$$\partial^k f = \begin{cases} f & \text{if } k = 0 \\ \partial(\partial^{k-1} f) & \text{if } k > 0 \end{cases}$$

The following is a basic, yet important result, given without proof.

Proposition 8.1. *If R_1, R_2 are rational relations on 2^*, then*

$$R_1 \circ R_2 = \{(u, v) \mid \exists w, uR_1w \text{ and } wR_2v\}$$

is a rational relation on 2^*.

The following lemmas help simplify the analysis of orbit rationality.

Lemma 8.5. *For $f \in \mathcal{S}(\mathcal{A})$, if f^k is orbit rational for some $k \in \mathbb{Z} \setminus \{0\}$, then f is orbit rational.*

Proof. Assume that k is a positive power of 2, since f and f^r are orbit equivalent for any odd r. Hence, it suffices to show that if f^2 is orbit rational, then f is orbit rational.
The relation (function) f is rational as it is implemented by $\mathcal{A}$. Since the relation $R_0 = (f^2)^*$ is rational by assumption, the relation

$$R_1 = \{(u, v) \mid v = uf^{2k+1} \text{ for some } k \in \mathbb{Z}\}$$

is rational by composition of relations. Then simply

$$f^* = R_0 \cup R_1$$

so it must be rational. □

Lemma 8.6. *For an even function $f \in \mathcal{S}(\mathcal{A})$, it is orbit rational if and only if ∂f is orbit rational.*

Proof. If f is orbit rational, By the quotient lemma, the orbit relation of ∂f is a quotient of the orbit relation of f, so it must be regular.
For the other direction note that

$$f^* = \{0 : 0, 1 : 1\} \cdot (\partial f)^* \cup \{\epsilon\}$$

and use the fact that $\partial f^k = (\partial f)^k$. □

The following is a partial converse to Lemma 8.5.

Lemma 8.7. *For an odd function $f \in \mathcal{S}(\mathcal{A})$, if it is orbit rational then f^n is orbit rational for any $n \in \mathbb{Z}$.*

Proof. Assume that $k > 0$. Then $n = 2^k \cdot l$ for some odd l. Since f^l is odd as well, it suffices to prove the claim for powers of 2 by induction.
Base case : The case when $k = 0$ is true by assumption.
Induction step : Assume that f^{2^k} is orbit rational. Then $\partial f^{2^{k+1}} = f^{2^k}$ so it follows from the Lemma 8.6. □

9 Conclusion

The study of invertible binary transducers is a fairly recent development and has significant consequences in both automata theory and group theory. This paper answers questions in both fields of study. Some decidability questions are answered using strictly automata theoretic techniques, while some automata structure theory questions are better explained by linear algebraic techniques.

10 Appendix

Below is the complete list of the characteristic polynomials of A matrices upto $m = 4$. There are 6 polynomials of degree 2, 14 polynomials of degree 3 and 36 polynomials of degree 4.

10.1 Degree 2

- $\chi_A(x) = x^2 - \frac{1}{2}$
- $\chi_A(x) = x^2 + \frac{1}{2}$
- $\chi_A(x) = x^2 - \frac{1}{2}x + \frac{1}{2}$
- $\chi_A(x) = x^2 + \frac{1}{2}x + \frac{1}{2}$
- $\chi_A(x) = x^2 - x + \frac{1}{2}$
- $\chi_A(x) = x^2 + x + \frac{1}{2}$

10.2 Degree 3

- $\chi_A(x) = x^3 + \frac{1}{2}x^2 - \frac{1}{2}x - \frac{1}{2}$
- $\chi_A(x) = x^3 - \frac{1}{2}$
- $\chi_A(x) = x^3 + \frac{1}{2}x^2 - \frac{1}{2}$
- $\chi_A(x) = x^3 + x^2 - \frac{1}{2}$
- $\chi_A(x) = x^3 - \frac{1}{2}x^2 + \frac{1}{2}x - \frac{1}{2}$
- $\chi_A(x) = x^3 + \frac{1}{2}x - \frac{1}{2}$
- $\chi_A(x) = x^3 - x^2 + x - \frac{1}{2}$
- $\chi_A(x) = x^3 - \frac{1}{2}x^2 - \frac{1}{2}x + \frac{1}{2}$
- $\chi_A(x) = x^3 - x^2 + \frac{1}{2}$
- $\chi_A(x) = x^3 - \frac{1}{2}x^2 + \frac{1}{2}$
- $\chi_A(x) = x^3 + \frac{1}{2}$
- $\chi_A(x) = x^3 + \frac{1}{2}x + \frac{1}{2}$
- $\chi_A(x) = x^3 + \frac{1}{2}x^2 + \frac{1}{2}x + \frac{1}{2}$
- $\chi_A(x) = x^3 + x^2 + x + \frac{1}{2}$

10.3 Degree 4

- $\chi_A(x) = x^4 + x^3 - x + \frac{1}{2}$
- $\chi_A(x) = x^4 + \frac{1}{2}x^3 - \frac{1}{2}x - \frac{1}{2}$
- $\chi_A(x) = x^4 - \frac{1}{2}$
- $\chi_A(x) = x^4 - \frac{1}{2}x^3 + \frac{1}{2}x^2 - \frac{1}{2}$
- $\chi_A(x) = x^4 + \frac{1}{2}x^3 + \frac{1}{2}x^2 - \frac{1}{2}$
- $\chi_A(x) = x^4 - \frac{1}{2}x^3 + \frac{1}{2}x - \frac{1}{2}$
- $\chi_A(x) = x^4 - x^3 + x - \frac{1}{2}$
- $\chi_A(x) = x^4 - x^3 + x^2 - x + \frac{1}{2}$
- $\chi_A(x) = x^4 - \frac{3}{2}x^3 + \frac{3}{2}x^2 - x + \frac{1}{2}$
- $\chi_A(x) = x^4 - \frac{1}{2}x^2 - \frac{1}{2}x + \frac{1}{2}$
- $\chi_A(x) = x^4 - \frac{1}{2}x^3 - \frac{1}{2}x + \frac{1}{2}$
- $\chi_A(x) = x^4 - \frac{1}{2}x + \frac{1}{2}$
- $\chi_A(x) = x^4 - x^3 + \frac{1}{2}x^2 - \frac{1}{2}x + \frac{1}{2}$
- $\chi_A(x) = x^4 - \frac{1}{2}x^3 + \frac{1}{2}x^2 - \frac{1}{2}x + \frac{1}{2}$
- $\chi_A(x) = x^4 - x^3 + x^2 - \frac{1}{2}x + \frac{1}{2}$
- $\chi_A(x) = x^4 - \frac{1}{2}x^3 + x^2 - \frac{1}{2}x + \frac{1}{2}$
- $\chi_A(x) = x^4 - x^2 + \frac{1}{2}$
- $\chi_A(x) = x^4 - \frac{1}{2}x^3 - \frac{1}{2}x^2 + \frac{1}{2}$
- $\chi_A(x) = x^4 - \frac{1}{2}x^2 + \frac{1}{2}$
- $\chi_A(x) = x^4 + \frac{1}{2}x^3 - \frac{1}{2}x^2 + \frac{1}{2}$
- $\chi_A(x) = x^4 - \frac{1}{2}x^3 + \frac{1}{2}$
- $\chi_A(x) = x^4 + \frac{1}{2}$
- $\chi_A(x) = x^4 + \frac{1}{2}x^3 + \frac{1}{2}$
- $\chi_A(x) = x^4 - \frac{1}{2}x^3 + \frac{1}{2}x^2 + \frac{1}{2}$
- $\chi_A(x) = x^4 + \frac{1}{2}x^2 + \frac{1}{2}$
- $\chi_A(x) = x^4 + \frac{1}{2}x^3 + \frac{1}{2}x^2 + \frac{1}{2}$
- $\chi_A(x) = x^4 + x^2 + \frac{1}{2}$
- $\chi_A(x) = x^4 - \frac{1}{2}x^2 + \frac{1}{2}x + \frac{1}{2}$
- $\chi_A(x) = x^4 + \frac{1}{2}x + \frac{1}{2}$
- $\chi_A(x) = x^4 + \frac{1}{2}x^3 + \frac{1}{2}x + \frac{1}{2}$
- $\chi_A(x) = x^4 + \frac{1}{2}x^3 + \frac{1}{2}x^2 + \frac{1}{2}x + \frac{1}{2}$

- $\chi_A(x) = x^4 + x^3 + \frac{1}{2}x^2 + \frac{1}{2}x + \frac{1}{2}$
- $\chi_A(x) = x^4 + \frac{1}{2}x^3 + x^2 + \frac{1}{2}x + \frac{1}{2}$
- $\chi_A(x) = x^4 + x^3 + x^2 + \frac{1}{2}x + \frac{1}{2}$
- $\chi_A(x) = x^4 + x^3 + x^2 + x + \frac{1}{2}$
- $\chi_A(x) = x^4 + \frac{3}{2}x^3 + \frac{3}{2}x^2 + x + \frac{1}{2}$

References

[1] Grigorchuk, R.I., "Degrees of growth of finitely generated groups and the theory of invariant means". Izvestiya Akademii Nauk SSSR. Seriya Matematicheskaya. vol. 48, no. 5, pp.939–985, 1984.

[2] Aleshin, S.V., "Finite automata and Burnside's problem for periodic groups", Mat. Zametki, 11:3, pp.319–328, 1972.

[3] Nekrashevych, V., Sidki, S., "Automorphisms of the binary tree: state-closed subgroups and dynamics of 1/2-endomorphisms," London Mathematical Society Lecture Note Series, Vol. 311, pp.375-404, 2004.

[4] Sutner, K., Lewi, K., "Iterating Binary Transducers," Journal of Automata, Language and Combinatorics *17*, pp.293-313, 2012.

[5] Nekrashevych, V., *Self-Similar Groups*, American Mathematical Society, 2005.

[6] Khoussainov, B., Nerode, A., *Automata Theory and its Applications*, Birkhäuser, 2001.

[7] Latimer, C.G., MacDuffee, C.C., "A Correspondence Between Classes of Ideals and Classes of Matrices," Annals of Mathematics, Second Series, Vol. 34 No. 2, pp.313-316, 1933.

[8] Newmann, M., *Integral Matrices*, Academic Press, New York, 1972.

[9] Bloch, A., Polya, G., "On the Roots of Certain Algebraic Equations", Proc. London Math. Soc 33 pp.102-114, 1932.