A TWO-RELATION MONOID DOES NOT HAVE FINITE DERIVATION TYPE

SLAVA PESTOV

ABSTRACT. The monoid presented by $\langle a, b \,|\, aaa = a, abba = bb\rangle$ does not have finite derivation type. Furthermore, a census of two-generator, two-relation monoids shows that this example is minimal with this property, if we order monoid presentations by the sum of lengths of their defining relations.

CONTENTS

1. INTRODUCTION

Imagine we list all monoid presentations with two generators and two relations, in order of increasing length N, and try to solve the word problem in each one:

$$\langle a, b \,|\, w = x, y = z\rangle \qquad N := |w| + |x| + |y| + |z|$$

How long before we find a monoid whose word problem cannot be solved by the Knuth-Bendix algorithm? The answer is, not long! The main result in Section 4 is that $\langle a, b \,|\, aaa = a, abba = bb\rangle$, of length 10, does not have *finite derivation type*, so Knuth-Bendix will always fail to find a finite complete presentation. Independent of the main result, Section 5 describes a computational survey of finite complete presentations for two-generator, two-relation monoids up to this length.

To backtrack a little, the *word problem* is perhaps the central question in the study of finitely-presented monoids:

> **The word problem.** Given a finite monoid presentation $\langle A \,|\, R\rangle$ and two words x, $y \in A^*$, can we rewrite x into y by a series of bidirectional rewrite steps taken from R?

The word problem is undecidable in the general case [1]. On the other hand, when $\langle A \,|\, R\rangle$ is a *complete* monoid presentation, we can compute a *normal form* for any word by viewing the elements of R as directed reduction rules. The word problem then reduces to checking if two words have identical normal forms [2]. The *Knuth-Bendix algorithm* takes an arbitrary monoid presentation, and attempts to construct a complete presentation by repeatedly adding new rules [3, 4, 5].

This process of *completion* either terminates in a finite number of steps, or "fails" in that it continues to record new consequences of the defining relations forever. It is known that a successful outcome can depend both on choice of reduction order, as well as the alphabet used to present the monoid [6, 7]. It is also known that there are finitely-presented monoids with no finite complete presentation over any alphabet, and we recall these results now.

A monoid with an undecidable word problem does not admit a finite complete presentation, so completion always fails to terminate in this case. Tseitin's classic example has 5 generators and 7 relations [8, 9]:

$$\begin{aligned}\mathfrak{C}_1 := \langle a, b, c, d, e \,|\, & ac = ca, ad = da, bc = cb, bd = db, \\ & eca = ce, edb = de, \\ & cca = ccae \rangle\end{aligned}$$

The first examples with decidable word problem are due to Craig C. Squier [10], who showed that a monoid given by a finite complete presentation must satisfy the invariant of FP_3, while S_k, with presentation below, has a decidable word problem for all $k \geq 0$, but not FP_3 when $k \geq 2$:

$$\begin{aligned}S_k := \langle a, b, t, x_1, \ldots, x_k, y_1, \ldots, y_k \,|\, & ab = 1, \\ & x_1 a = a t x_1, \quad \ldots, \quad x_k a = a t x_k, \\ & x_1 t = t x_1, \quad \ldots, \quad x_k t = t x_k, \\ & x_1 b = b x_1, \quad \ldots, \quad x_k b = b x_k, \\ & x_1 y_1 = 1, \quad \ldots, \quad x_k y_k = 1 \rangle\end{aligned}$$

To settle the $k = 1$ case, Squier then introduced the invariant of *finite derivation type*, or FDT, in a subsequent paper [11]. The key result is that a monoid given by a finite complete presentation has FDT, while S_1, with 5 generators and 5 relations, does not have FDT, and thus, no finite complete presentation:

$$S_1 := \langle a, b, t, x, y \,|\, ab = 1, xa = atx, xt = tx, xb = bx, xy = 1 \rangle$$

We review the definition of FDT in Section 3. We do not define FP_3; it suffices to note that FDT implies FP_3, so in fact S_k is not FDT for all $k \geq 1$ [12].

Finite derivation type is not a *sufficient* condition for a monoid to admit a finite complete presentation. Katsura and Kobayashi discovered this monoid with decidable word problem and FDT, but no finite complete presentation [13]:

$$\begin{aligned}\langle a, b_1, c_1, d_1, b_2, c_2, d_2, b_3, c_3, d_3 \,|\, & b_1 a = a b_1, b_2 a = a b_2, b_3 a = a b_3, \\ & c_1 b_1 = c_1 b_1, c_2 b_2 = c_1 b_1, \\ & b_1 d_1 = b_1 d_1, b_2 d_2 = b_1 d_1 \rangle\end{aligned}$$

Can we have fewer defining relations than Squier's S_1? It seems the previous record was *three* relations. Lafont and Prouté exhibit this non-FP_3 monoid [14]:

$$\langle a, b, c, d, d' \,|\, ab = a, da = ac, d'a = ac \rangle$$

Cain et al. show that this monoid does not have FDT [15]:

$$\langle a, b, c \,|\, ac = ca, bc = cb, cab = cbb \rangle$$

Every one-relation monoid $\langle A \,|\, u = v \rangle$ has FDT [16]. It remains an open question if every one-relation monoid has a finite complete presentation or a decidable word problem [17].

2. Preliminaries

This article assumes some familiarity with finitely-presented monoids and string rewriting; this section is too dense to serve as a proper introduction to the topic. A good reference can be found in [18].

Definition 2.1. A *monoid* is a set M with an associative binary operation $\cdot_M$ and identity element 1_M. A *monoid homomorphism* $\phi\colon M \to N$ satisfies $\phi(1_M) = 1_N$, and $\phi(x \cdot_M y) = \phi(x) \cdot_N \phi(y)$ for all $x, y \in A^*$.

Definition 2.2. If A is any set, the *free monoid* A^* is the set of all finite sequences of elements of A. For our purposes, the alphabet A is always finite.

- The *length* of $x \in A^*$ is denoted by $|x| \geq 0$.
- An $x \in A$ is a *letter* and an $x \in A^*$ is a *word*.
- The unique *empty word* of length 0 is denoted by 1.
- We view each letter $x \in A$ as a word of length 1 in A^*.
- The concatenation of words x and y is denoted xy or $x \cdot y$.
- If $x \in A^*$, then x^n means x concatenated with itself n times.
- A word u is a *factor* of a word w if $w = xuy$ for some $x, y \in A^*$.
- The equality operator $=$ denotes graphical equality of words in A^*.

A *monoid presentation* is a pair $\langle A \mid R \rangle$, where A is a set, and $R \subset A^* \times A^*$ is a set of ordered pairs of words. A presentation is *finite* if A and R are finite sets.

- A *rewrite step* (over R) is a quadruple $x \cdot (u, v) \cdot y$, where $x, y \in A^*$ are called the left and right *whiskers*, and either (u, v) or $(v, u) \in R$. The *source* of this rewrite step is the word xuy, and the *destination* is xvy.
- A *rewrite path* is either an empty rewrite path 1_w for a word $w \in A^*$, or the composition of one or more rewrite steps $p = s_1 \triangleright \cdots \triangleright s_n$ where the source of each step is identical to the destination of the previous step.
- The *monoid congruence* $\Leftrightarrow_R$ relates all pairs $x, y \in A^*$ such that there is a rewrite path from x to y.
- The equivalence class of $x \in A^*$ is denoted by $[\![x]\!]_R$. The equivalence classes of $\Leftrightarrow_R$ then have the structure of a monoid, with identity element $[\![1]\!]_R$ and binary operation $[\![x]\!]_R \cdot [\![y]\!]_R := [\![x \cdot y]\!]_R$.
- A *finitely-presented monoid* is one that is isomorphic to a monoid obtained in this way, from a finite presentation.

We also need to consider rewriting steps that only apply a rule from left to right:

- A rewrite step $x \cdot (u, v) \cdot y$ is *positive* if $(u, v) \in R$, and *negative* if $(v, u) \in R$.
- A rewrite path is *positive* if every step taken is positive.
- The *reduction relation* $\Rightarrow_R$ relates all pairs $x \Rightarrow_R y$ such that there is a positive rewrite path from x to y.
- The reduction relation $\Rightarrow_R$ is *terminating* if there is no infinite sequence of positive rewrite steps where the source of each step is identical to the destination of the previous step.
- The reduction relation $\Rightarrow_R$ is *confluent* if whenever $x \Rightarrow_R y$ and $x \Rightarrow_R z$, there exists a word $w \in A^*$ such that $y \Rightarrow_R w$ and $z \Rightarrow_R w$.
- A word $x \in A^*$ is *irreducible* if $x \Rightarrow_R y$ implies that $x = y$.
- If y is irreducible and $x \Rightarrow_R y$, we say that y is a *normal form* for x.

A monoid presentation $\langle A \,|\, R\rangle$ is *complete* if $\Rightarrow_R$ is terminating and confluent. A *finite complete presentation* is one that is both finite, and complete. In this case, every equivalence class of $\Leftrightarrow_R$ has an effectively computable, unique normal form.

Knuth-Bendix. A sufficient condition for the termination of a reduction relation $\Rightarrow$ is that for each $(u, v) \in R$, we have $u > v$ for some suitable reduction order on A^*:

Definition 2.3. A *reduction order* on A^* is a well-founded linear order that is *closed under translation*, so $u > v$ implies that $xuy > xvy$ for all u, v, x, $y \in A^*$.

The most important reduction order for our purposes is the *shortlex order*. We take a linear order on the alphabet A and extend it to pairs of words x, $y \in A^*$ as follows. If $|x| < |y|$, then $x < y$; otherwise if $|x| = |y|$, we compare the letters of x and y from left to right.

Finally, to establish the confluence of a terminating reduction relation, we will need Newman's lemma [19], which we recall after some preliminary definitions.

Definition 2.4. Two words u, $u' \in A^*$ *overlap* if one of the following is true:

(1) The first has a suffix equal to a prefix of the second, so $u = xy$ and $u' = yz$, for some words x, y, z, with the overlapping piece y non-empty.
(2) The second contains the first as a factor, so $u' = xuy$, for some words x, y.

Definition 2.5. Let $\langle A \,|\, R\rangle$ be a monoid presentation. Two rules $(u, v) \in R$ and $(u', v') \in R$ *overlap* if their left-hand sides overlap as above.

- In our examples, only overlaps of the first kind will occur. (In the second case, we can always remove one of the rules from the presentation.)
- The two ways of reducing the overlap word give us a pair of positive rewrite steps, called a *critical pair*.
- An important fact is that the left-hand sides of two rules may overlap more than once, at different positions; this yields distinct critical pairs.
- A critical pair is *trivial* if both sides can be reduced to the same word by applying the normal form algorithm.

Lemma 2.6. A terminating reduction relation $\Rightarrow$ is confluent if and only if all critical pairs are trivial.

Thus, we can use this result to show that a terminating reduction relation is confluent by considering all overlaps among the rules, and for each one, exhibiting a pair of positive rewrite paths (l, r), where the source of each path is the overlap word, and the destination is any common descendant of the two sides.

The above also forms the basis for the Knuth-Bendix algorithm [3, 4, 5]. We check for confluence first, and repair any violations we find by adding new rules. This can introduce new confluence violations, so the process repeats, possibly forever.

3. Finite Derivation Type

We now attempt to summarize Squier's paper [11]. The essential idea is this. A rewrite path from x to y is a proof that $x \Leftrightarrow y$, and there may be more than one such path, if there are "multiple ways" of rewriting x into y. We say that two rewrite paths are *parallel* if they have the same source and same destination, and a presentation has *finite derivation type* if this parallelism relation, which describes "all the ways" of rewriting with those rules, is finitely generated. Finally, in the case of a complete presentation, we can explicitly describe this generating set.

The *derivation graph* of a monoid presentation $\langle A \mid R \rangle$, denoted by $\Gamma(A \mid R)$, is a directed graph whose vertices are words in A^*, and edges are rewrite steps formed from the defining relations of R. Our next goal is to develop a "two-dimensional algebra of rewrite paths" on this graph.

- A rewrite path as defined earlier is a path in $\Gamma(A \mid R)$, and an equivalence class of the monoid congruence $\Leftrightarrow_R$ is a connected component of $\Gamma(A \mid R)$.
- The set of all paths in $\Gamma(A \mid R)$ is denoted by $P(A \mid R)$.
- The *length* of a rewrite path is the number of steps taken.
- The *composition* of p_1 and p_2, where $\mathsf{src}(p_2) = \mathsf{dst}(p_1)$, performs the steps of p_1 followed by p_2. The resulting path is denoted by $p_1 \triangleright p_2$.
- The *inverse* of a rewrite step $x \cdot (u, v) \cdot y$ is $x \cdot (v, u) \cdot y$. The inverse p^{-1} of a rewrite path p performs the inverse of each step of p in reverse order.
- The left and right *whiskering actions* of A^* on $P(A \mid R)$ extend the left or right whiskers of each step in a path by concatenation with a word $z \in A^*$:

$$z \cdot x \cdot (u, v) \cdot y := zx \cdot (u, v) \cdot y$$
$$x \cdot (u, v) \cdot y \cdot z := x \cdot (u, v) \cdot yz$$

- Whiskering $\cdot$ has higher precedence (binds tighter) than composition $\triangleright$.
- The parallelism relation is denoted by $\|$. Recall that it relates all p, q such that $\mathsf{src}(p) = \mathsf{src}(q)$ and $\mathsf{dst}(p) = \mathsf{dst}(q)$.
- The parallelism relation $\|$ is an equivalence relation on $P(A \mid R)$.

Definition 3.1. Let $\langle A \mid R \rangle$ be a monoid presentation, and let $\approx$ be an equivalence relation on $P(A \mid R)$. We say that $\approx$ is a *homotopy relation* on $P(A \mid R)$ if it satisfies the following properties:

(1) (Parallelism.) If $p \approx q$, then:
$$p \parallel q$$
(2) (Composition.) If $p \approx q$, $\mathsf{dst}(r) = \mathsf{src}(p)$, and $\mathsf{dst}(p) = \mathsf{src}(s)$, then:
$$r \triangleright p \triangleright s \approx r \triangleright q \triangleright s$$
(3) (Whiskering.) If $p \approx q$, and x, $y \in A^*$, then:
$$x \cdot p \cdot y \approx x \cdot q \cdot y$$
(4) (Complementary steps.) If s is a rewrite step, then:
$$s \triangleright s^{-1} \approx 1_{\mathsf{src}(s)}$$
(5) (Interchange of disjoint steps.) If s_1 and s_2 are two rewrite steps, then:
$$s_1 \cdot \mathsf{src}(s_2) \triangleright \mathsf{dst}(s_1) \cdot s_2 \approx \mathsf{src}(s_1) \cdot s_2 \triangleright s_1 \cdot \mathsf{dst}(s_2)$$

Some immediate consequences of the above:

(1) The parallelism relation $\|$ is a homotopy relation on $P(A \mid R)$, and it is the *largest* homotopy relation, in the sense that every homotopy relation on $P(A \mid R)$ is contained in $\|$.
(2) The set of all homotopy relations on $P(A \mid R)$ is closed under arbitrary intersection and directed union.

From (1) and (2), it follows that if B is any subset of $\|$, there is a unique smallest homotopy relation on $P(A \mid R)$ that contains B. This is the homotopy relation *generated* by B, and we denote it by $\approx_B$ below.

Just as a monoid congruence can be understood in terms of rewrite paths between words, we can show that the generated homotopy relation $\approx_B$ relates all pairs of paths where the first path can be obtained from the second path by a finite sequence of two-dimensional *elementary transformations*, described in the following lemma. This is Theorem 3.4 in [11], or Proposition 2.4 in [20].

Lemma 3.2. Let $\langle A \mid R \rangle$ be a monoid presentation, let $B \subset \|$, and suppose we have two paths p, $q \in P(A \mid R)$. We have $p \approx_B q$ if and only if there exist a sequence of paths $p_1, \ldots, p_n$ such that $p = p_1$, $q = p_n$, and each p_{i+1} is obtained from p_i by one of the below elementary transformations:

(1) Removal of complementary steps $s \triangleright s^{-1}$:
$$p_i = p_i' \triangleright (s \triangleright s^{-1}) \triangleright p_i''$$
$$p_{i+1} = p_i' \triangleright p_i''$$
(2) Insertion of complementary steps $s \triangleright s^{-1}$.
(3) Interchange of two disjoint steps $\mathsf{src}(s_1) \cdot s_2$ and $s_1 \cdot \mathsf{dst}(s_2)$, so the leftmost rewrite is now performed first:
$$p_i = p_i' \triangleright (\mathsf{src}(s_1) \cdot s_2 \triangleright s_1 \cdot \mathsf{dst}(s_2)) \triangleright p_i''$$
$$p_{i+1} = p_i' \triangleright (s_1 \cdot \mathsf{src}(s_2) \triangleright \mathsf{dst}(s_1) \cdot s_2) \triangleright p_i''$$
(4) Interchange of two disjoint steps in the other direction.
(5) Positive replacement from B, so for some $(l, r) \in B$, and x, $y \in A^*$, and paths p_i', p_i'' satisfying the necessary conditions on source and destination:
$$p_i = p_i' \triangleright (x \cdot l \cdot y) \triangleright p_i''$$
$$p_{i+1} = p_i' \triangleright (x \cdot r \cdot y) \triangleright p_i''$$
(6) Negative replacement from B, so as above but with p_i and p_{i+1} swapped.

Proof. Let's say that $\equiv_B$ is the equivalence relation on $P(A \mid R)$ that relates all pairs of paths that can be joined by a sequence of elementary transformations, listed above. We argue that $\equiv_B$ is the same relation as $\approx_B$.

Suppose that $p \equiv_B q$. The source of each elementary transformation is parallel to the destination, and an induction on the number of elementary transformations from p to q shows that $p \parallel q$. Furthermore, the source and destination of each elementary transformation is actually equivalent under $\approx_B$. So $\equiv_B \subseteq \approx_B$ as a direct consequence of Definition 3.1. Thus, $p \approx_B q$. In other words, $\equiv_B \subseteq \approx_B$.

Now, suppose that $(p, q) \in B$. There is an elementary transformation of type (5) with source p and destination q, so $p \equiv_B q$. Thus, $B \subseteq \equiv_B$. Furthermore, it is easy to see that $\equiv_B$ itself satisfies the conditions of a homotopy relation. Together these two facts imply that $\approx_B \subseteq \equiv_B$. □

When $B = \varnothing$ in Lemma 3.2, we get elementary transformations of the first four kinds only, so two paths are equivalent under $\approx_\varnothing$ if each one can be obtained from the other by some combination of inserting and removing complementary rewrite steps, and interchanging disjoint rewrite steps.

Definition 3.3. The *null homotopy relation* is the homotopy relation generated by the empty set. Instead of writing $\approx_\varnothing$, we denote it by $\simeq$.

Note that $\simeq$ is the *smallest* homotopy relation, in the sense that every homotopy relation contains $\simeq$. We can now define the key concept of this section.

Definition 3.4. A monoid presentation $\langle A \mid R \rangle$ has *finite derivation type* if some finite subset $B \subset \parallel$ generates $\parallel$ as a homotopy relation on $P(A \mid R)$.

We summarize our notation for homotopy relations before we continue:

$\approx$	Some homotopy relation.
$\simeq$	The null homotopy relation.
$\approx_B$	Homotopy relation generated by B.
$\parallel$	The parallelism homotopy relation.

Circuits. Under the preceding definitions, a homotopy relation is a generated by a set of pairs of parallel paths. However, it is sometimes more convenient to work with a set of *circuits* in the derivation graph instead:

- A rewrite path $c \in P(A \mid R)$ is a *circuit* if $\mathsf{src}(c) = \mathsf{dst}(c)$.
- We say $\mathsf{src}(c)$ is the *basepoint* of the circuit.
- The set of all circuits in $\Gamma(A \mid R)$ is denoted by $C(A \mid R)$.

Every homotopy relation can be generated by a set of circuits. In the below, I is some arbitrary index set.

Definition 3.5. If $\mathcal{B} := \{c_i\}_{i \in I}$ is a subset of $C(A \mid R)$, let $B := \{(c_i, 1_{\mathsf{src}(c_i)})\}_{i \in I}$. Note that $B \subset \parallel$. The homotopy relation *generated* by $\mathcal{B}$ is defined as the homotopy relation generated by B in the earlier sense.

Lemma 3.6. If $B := \{(p_i, q_i)\}_{i \in I}$ is a subset of $\parallel$, then $\mathcal{B} := \{q_i \triangleright p_i^{-1}\}_{i \in I}$ generates the same homotopy relation as B.

Proof. First, suppose that $c \in \mathcal{B}$. We will show that $c \approx_B 1_{\mathsf{src}(c)}$. By definition, $c = q_i \triangleright p_i^{-1}$ for some $i \in I$, so $(p, q) \in B$. Since $p \approx_B q$ and $\approx_B$ is a homotopy relation, we also have $p \triangleright p^{-1} \approx_B q \triangleright p^{-1}$, or in other words, $c \approx_B 1_{\mathsf{src}(c)}$.

For the other direction, suppose that $(p, q) \in B$. To see that $p \approx_{\mathcal{B}} q$, notice that $p \triangleright q^{-1} \approx_{\mathcal{B}} 1_{\mathsf{src}(p)}$, and compose both sides on the right with q. □

We give a name to those sets of circuits that generate the parallelism relation $\parallel$ as a homotopy relation.

Definition 3.7. Let $\langle A \mid R \rangle$ be a monoid presentation. A subset $\mathcal{B} \subset C(A \mid R)$ is a *homotopy base* for $\langle A \mid R \rangle$ if $\mathcal{B}$ generates $\parallel$ as a homotopy relation.

The next lemma reformulates Lemma 3.2 in terms of circuits. The idea is that up to null homotopy, an arbitrary circuit $c \in C(A \mid R)$ can be constructed by "gluing" together a combination of the circuits in a homotopy base, suitably whiskered. This appears as Lemma 2.1 in [16], or Proposition 2.2 in [20].

Lemma 3.8. Let $\langle A \mid R \rangle$ be a monoid presentation, and let $\mathcal{B} \subset C(A \mid R)$ be a set of circuits. Then $\mathcal{B}$ is a homotopy base if and only if for any circuit $c \in C(A \mid R)$, there exist a series of:

(1) words x_i, $y_i \in A^*$,
(2) paths $p_i \in P(A \mid R)$,
(3) circuits $q_i \in \mathcal{B}$,
(4) and integers $e_i \in \{-1, 1\}$,

satisfying:

$$c \simeq p_1 \triangleright (x_1 \cdot q_1^{e_1} \cdot y_1) \triangleright p_1^{-1} \triangleright \cdots \triangleright p_n \triangleright (x_n \cdot q_n^{e_n} \cdot y_n) \triangleright p_n^{-1}$$

with $\mathsf{src}(p_i) = \mathsf{src}(c)$ and $\mathsf{dst}(p_i) = x_i \cdot \mathsf{src}(q_i) \cdot y_i$ for each $i = 1, \ldots, n \geq 0$.

Proof. ("If") Starting from the assumption that every $c \in C(A \mid R)$ can be expressed as a combination of circuits in $\mathcal{B}$, we can show that $\mathcal{B}$ is a homotopy base as follows. Take arbitrary p, $q \in P(A \mid R)$ such that $p \parallel q$, and let $c := q \triangleright p^{-1}$. Since $c \in C(A \mid R)$, we can write it as a combination of circuits from $\mathcal{B}$.

For each term $x_i \cdot q_i^{e_1} \cdot y_i$ appearing in the above expression for c, we have $q_i \in \mathcal{B}$, and thus $x_i \cdot q_i^{e_1} \cdot y_i \approx_{\mathcal{B}} 1_{\mathsf{dst}\, p_i}$. After contracting each such term, we see the remaining p_i and p_i^{-1} terms then cancel out, so $c \approx_{\mathcal{B}} 1_{\mathsf{src}(c)}$. Composing both sides on the right by p, we get $q \approx_B p$.

("Only if") Suppose that $\mathcal{B}$ is a homotopy base, and let $B := \{(c_i, 1_{\mathsf{src}(c_i)})\}$ be the corresponding set of pairs of parallel paths. Suppose we are given an arbitrary $c \in C(A \mid R)$. Since $\mathcal{B}$ is a homotopy base, we have $c \approx_B 1_{\mathsf{src}(c)}$. By Lemma 3.2, there exists a series of elementary transformations from c to $1_{\mathsf{src}(c)}$, using B.

We perform an induction on the number of transformations. In the base case, $c = 1_{\mathsf{src}(c)}$, and we're done. Otherwise, we can assume that c can be transformed into c' by a single elementary transformation, while c' can be transformed into $1_{\mathsf{src}(c)}$ in one fewer step than c.

By induction, the result already holds for c', so we can write c' as a combination of whiskered circuits from $\mathcal{B}$:

$$c' \simeq p_1 \triangleright (x_1 \cdot q_1^{e_1} \cdot y_1) \triangleright p_1^{-1} \triangleright \cdots \triangleright p_n \triangleright (x_n \cdot q_n^{e_n} \cdot y_n) \triangleright p_n^{-1}$$

We consider each kind of elementary transformation from c' to c:

(1) If c is obtained from c' by inserting or removing complementary rewrite steps, or interchanging disjoint rewrite steps, then $c' \simeq c$, so we're done.
(2) Otherwise, for some $(l, r) \in B$ or $(r, l) \in B$, we have:

$$c' = s \triangleright (x \cdot l \cdot y) \triangleright t$$
$$c \;\; = s \triangleright (x \cdot r \cdot y) \triangleright t$$

By definition of B, one of l or r is an element of $\mathcal{B}$, while the other one is the empty path $1_{\mathsf{src}(l)}$. If $l \in \mathcal{B}$, then $c = s \triangleright t$, and so:

$$c \simeq c' \triangleright t^{-1} \triangleright (x \cdot l^{-1} \cdot y) \triangleright t$$

If $r \in \mathcal{B}$, then $c' = s \triangleright t$, and so:

$$c \simeq c' \triangleright t^{-1} \triangleright (x \cdot r^{-1} \cdot y) \triangleright t$$

In every case, c again has the required form. This closes the induction. □

With the preliminaries out of the way, we state the two key results of Squier's paper without proof.

The first is that the property of finite derivation type is an invariant of the presented monoid, independent of choice of finite presentation. We don't actually need this fact, but we recall it for posterity. This is Theorem 4.3 in [11].

Lemma 3.9. Suppose that $\langle A \mid R \rangle$ and $\langle C \mid S \rangle$ are two finite presentations that present isomorphic monoids. Then $\langle A \mid R \rangle$ has finite derivation type if and only if $\langle C \mid S \rangle$ has finite derivation type.

Sketch of Proof. We can obtain $\langle A \mid R \rangle$ from $\langle C \mid S \rangle$ by a finite sequence of Tietze transformations (adding or removing a generator equivalent to a word, adding or removing a rule given by a rewrite path over the remaining rules). For each kind of elementary Tietze transformation, one can translate a finite homotopy base for the source presentation into a finite homotopy base for the destination presentation. □

The second key fact we state without proof is the connection between complete presentations and homotopy relations. This fact, we do need. For the proof, see Theorem 5.2 in [11].

Definition 3.10. Let $\langle A \mid R \rangle$ be a complete presentation. The set of *critical circuits* in $\Gamma(A \mid R)$ is the set of all rewrite paths of the form $r \triangleright l^{-1}$, as (l, r) ranges over the pairs of positive rewrite paths resolving each critical pair.

Lemma 3.11. If $\langle A \mid R \rangle$ is a complete, the set of all critical circuits in $\Gamma(A \mid R)$ is a homotopy base for $\langle A \mid R \rangle$.

The above is true even when R is infinite; we get an infinite homotopy base in that case. However, when R also happens to be finite, we get Theorem 5.3 in [11]:

Lemma 3.12. Suppose that $\langle A \mid R \rangle$ is finite and complete. Then the presented monoid has finite derivation type.

Proof. Since $\langle A \mid R \rangle$ is complete, the set of critical circuits in $\Gamma(A \mid R)$ forms a homotopy base by Lemma 3.11, and since R is finite, this set of circuits is finite. □

Thus, if we can show that *some* finite presentation of a monoid does not have finite derivation type, we can conclude that this monoid does not have a finite complete presentation.

Mapping of rewrite paths. We need two more technical results. The first allows us to translate the critical circuits of an infinite complete presentation into a homotopy base for a finite, but not necessarily complete, presentation of the same monoid.

In Squier's paper, the below is a consequence of Theorem 3.6, Corollary 3.7, and Lemma 4.4. Since we skip the details of those proofs, we give a self-contained argument for our special case.

Definition 3.13. Let $\langle A \mid \tilde{R} \rangle$ and $\langle A \mid R \rangle$ be any two monoid presentations, sharing the same alphabet. A function $\Phi \colon P(A \mid \tilde{R}) \to P(A \mid R)$ is a *mapping of rewrite paths* if it satisfies the following conditions:

(1) (Compatibility.) For all $p \in P(A \mid \tilde{R})$:

$$\mathsf{src}(\Phi(p)) = \mathsf{src}(p)$$
$$\mathsf{dst}(\Phi(p)) = \mathsf{dst}(p)$$

(2) (Composition.) For all $p, q \in P(A \mid \tilde{R})$:

$$\Phi(p \triangleright q) = \Phi(p) \triangleright \Phi(q)$$

(3) (Whiskering.) For all $p \in P(A \mid \tilde{R})$ and $x, y \in A^*$:

$$\Phi(x \cdot p \cdot q) = x \cdot \Phi(p) \cdot q$$

(4) (Inverses.) For all $p \in P(A \mid \tilde{R})$:

$$\Phi(p^{-1}) = \Phi(p)^{-1}$$

(5) (Identity.) For all $x \in A^*$:

$$\Phi(1_x) = 1_x$$

Notice how a mapping of rewrite paths is completely determined by the image of each $(u, v) \in \tilde{R}$, viewed as a positive rewrite step with empty whiskers.

The above describes a more general situation than we need. We will impose two additional conditions on $\langle A \,|\, \tilde{R} \rangle$ and $\langle A \,|\, R \rangle$:

(1) We require that $\tilde{R}$ and R generate the same monoid congruence on A^*.
(2) Also, R must be a subset of $\tilde{R}$.

Under these conditions, for each $(u, v) \in \tilde{R} \setminus R$, there exists a rewrite path that we denote by $p_{u,v} \in P(A \,|\, R)$, with $\mathsf{src}(p_{u,v}) = u$ and $\mathsf{dst}(p_{u,v}) = v$.

From these paths $p_{u,v}$, we can define a mapping $\Phi \colon P(A \,|\, \tilde{R}) \to P(A \,|\, R)$ which maps each element of R to itself, and each $(u, v) \in \tilde{R} \setminus R$ to the corresponding path $p_{u,v}$. We can use this mapping to transform a homotopy base for $\langle A \,|\, \tilde{R} \rangle$ into a homotopy base for $\langle A \,|\, R \rangle$:

Lemma 3.14. Let $\Phi \colon P(A \,|\, \tilde{R}) \to P(A \,|\, R)$ be as above. If some set $\tilde{B}$ generates $\parallel$ as a homotopy relation on $P(A \,|\, \tilde{R})$, then $\Phi(\tilde{B})$ generates $\parallel$ as a homotopy relation on $P(A \,|\, R)$, where we define $\Phi(\tilde{B})$ as the set of pairs of paths $(\Phi(l), \Phi(r))$ obtained by applying Φ to both sides of each $(l, r) \in \tilde{B}$.

Proof. Under the conditions, $P(A \,|\, R) \subseteq P(A \,|\, \tilde{R})$, and Φ is the identity mapping on this subset $P(A \,|\, R)$. Let $B := \Phi(\tilde{B})$.

Suppose that $p \parallel q$ for some $p, q \in P(A \,|\, R)$. We must show that $q \approx_B q$. Since $\tilde{B}$ generates $\parallel$ as a homotopy relation on $P(A \,|\, \tilde{R})$, we have $p \approx_{\tilde{B}} q$. By Lemma 3.2, we can exhibit a sequence of elementary transformations involving $\tilde{B}$, transforming p into q. By induction, it suffices to show that in the case of a single elementary transformation, we have $p \approx_B q$. We consider each kind of elementary transformation in turn:

(1) If p is obtained from q by inserting or removing complementary rewrite steps, or by interchanging disjoint rewrite steps, then we actually have $p \simeq q$, and so in particular, $p \approx_B q$.
(2) Otherwise, p is obtained from q by a positive or negative replacement from $\tilde{B}$. So, for some (l, r) or $(r, l) \in \tilde{B}$, and $x, y \in A^*$:

$$p = p' \triangleright (x \cdot l \cdot y) \triangleright p''$$
$$q = q' \triangleright (x \cdot r \cdot y) \triangleright q''$$

We have either $(\Phi(l), \Phi(r)) \in B$ or $(\Phi(l), \Phi(r)) \in B$ from the definition of B, and since $\approx_B$ is a homotopy relation:

$$p' \triangleright (x \cdot l \cdot y) \triangleright p'' \approx_B q' \triangleright (x \cdot r \cdot y) \triangleright q''$$

We see that $p \approx_B q$ for each kind of elementary transformation, and by induction, $p \approx_B q$ for arbitrary p and q such that $p \parallel q$. Thus, B generates $\parallel$ on $P(A \,|\, R)$. □

If we start with a set of circuits $\tilde{\mathcal{B}}$ and apply the above lemma to the set $\tilde{B} \subseteq \parallel$ formed from $\tilde{\mathcal{B}}$ as in Definition 3.5, we get the same thing in terms of circuits:

Lemma 3.15. Let $\Phi \colon P(A \,|\, \tilde{R}) \to P(A \,|\, R)$ be as above. If $\tilde{\mathcal{B}} \subseteq C(A \,|\, \tilde{R})$ is a homotopy base for $\langle A \,|\, \tilde{R} \rangle$, then $\Phi(\tilde{\mathcal{B}}) \subseteq C(A \,|\, R)$ is a homotopy base for $\langle A \,|\, R \rangle$.

Our final preliminary lemma states that if a monoid presentation has finite derivation type, then every homotopy base for this presentation already *contains* a finite homotopy base. Thus, to show that a monoid does *not* have finite derivation type, it suffices to exhibit a finite presentation, with an infinite homotopy base, no finite subset of which is a homotopy base.

Lemma 3.16. Let $\langle A \,|\, R\rangle$ be a monoid presentation, and let $\mathcal{B}$ and $\mathcal{C}$ be two subsets of $C(A \,|\, R)$. If $\mathcal{B}$ is a finite homotopy base for $\langle A \,|\, R\rangle$, and $\mathcal{C}$ is any homotopy base for $\langle A \,|\, R\rangle$, then there is a a finite subset $\mathcal{C}' \subset \mathcal{C}$ such that $\mathcal{C}'$ is itself a homotopy base for $\langle A \,|\, R\rangle$.

Proof. Since $\mathcal{C}$ is a homotopy base, we can express each circuit of $\mathcal{B}$ in terms of $\mathcal{C}$, in the sense of Lemma 3.8. Since $\mathcal{B}$ is finite, we only need a finite subset $\mathcal{C}' \subset \mathcal{C}$ to express everything in $\mathcal{B}$. But $\mathcal{B}$ is a homotopy base, so every $c \in C(A \,|\, R)$ can be expressed in terms of the circuits of $\mathcal{C}'$, and so $\mathcal{C}'$ is a homotopy base. □

4. The Main Result

Theorem 1. The monoid $M := \langle a, b \,|\, aaa = a, abba = bb\rangle$ does not have finite derivation type, and thus, no finite complete presentation.

The proof takes the rest of this section. Here is a summary:

(1) We introduce a longer presentation $\langle A \,|\, R\rangle$ of M that is easier to work with.
(2) We extend this presentation $\langle A \,|\, R\rangle$ to an *infinite* complete presentation $\langle A \,|\, R^\infty\rangle$. An aside will establish that M has a decidable word problem.
(3) Considering the overlaps between the left-hand sides of R^∞ yields a set of critical circuits in $\Gamma(A \,|\, R^\infty)$. By Lemma 3.11, this infinite set of circuits is a homotopy base for $\langle A \,|\, R^\infty\rangle$.
(4) We define a mapping of rewrite paths $\Phi\colon P(A \,|\, R^\infty) \to P(A \,|\, R)$, which sends each rewrite rule in $R^\infty \setminus R$ to a fixed rewrite path over R.
(5) We apply Φ to each critical circuit in $\Gamma(A \,|\, R^\infty)$. By Lemma 3.15, this set of circuits, that we call $\mathcal{C}$, is a homotopy base for $\langle A \,|\, R\rangle$.
(6) We minimize $\mathcal{C}$ somewhat, to get a homotopy base $\mathcal{B} \subset \mathcal{C}$. We then show that no finite subset $\mathcal{B}' \subset \mathcal{B}$ is a homotopy base for $\langle A \,|\, R\rangle$.
(7) We conclude that M does not have finite derivation type, by Lemma 3.16.

Let $A := \{a, b\}$ and $R' := \{(aaa, a), (abba, bb)\}$ denote the alphabet and defining relations of our original presentation of M. To construct our infinite complete presentation $\langle A \,|\, R^\infty\rangle$, we start from this presentation of M instead:

$$\langle a, b \,|\, aaa = a, bba = abb, aabb = bb\rangle$$

Henceforth, this set of three rules above will be called $R := \{\alpha, \beta, \gamma_0\}$, where:

$$\begin{aligned} \alpha &:= (aaa, a) \\ \beta &:= (bba, abb) \\ \gamma_0 &:= (aabb, bb) \end{aligned}$$

Our first task is to prove that $\langle A \,|\, R\rangle$ and $\langle A \,|\, R'\rangle$ present the same monoid.

Lemma 4.1. Both R and R' generate the same monoid congruence on A^*.

Proof. First, we show that $(abba, bb) \in R'$ belongs to $\Leftrightarrow_R$. We have $abba \Leftrightarrow_R bb$, because:

$$abba \Leftrightarrow_R aabb \Leftrightarrow_R bb$$

Next, we're going to show that β and $\gamma_0 \in R$ belong to $\Leftrightarrow_{R'}$. We have $bba \Leftrightarrow_{R'} abb$:

$$bba \Leftrightarrow_{R'} abbaa_{R'} \Leftrightarrow aaabbaa \Leftrightarrow_{R'} aabba \Leftrightarrow_{R'} abb$$

and $aabb \Leftrightarrow_{R'} bb$:

$$aabb \Leftrightarrow_{R'} aaabba \Leftrightarrow_{R'} abba \Leftrightarrow_{R'} bb$$

Thus, $\Leftrightarrow_R = \Leftrightarrow_{R'}$ as sets, and the mapping that sends each letter to itself induces a monoid isomorphism between the monoids presented by $\langle A \,|\, R\rangle$ and $\langle A \,|\, R'\rangle$. $\square$

In the rest of this section, this monoid congruence will be denoted by $\Leftrightarrow$.

Notice how rule β says that bb commutes with a, and since bb always commutes with b, we actually have $xbb \Leftrightarrow bbx$ for all $x \in A^*$. In particular, $(ba)^n bb \Leftrightarrow bb(ba)^n$ for all $n \geq 0$. Since also $aabb \Leftrightarrow bb$ via γ_0, we have, for all $n \geq 0$:

$$aa(ba)^n bb \Leftrightarrow aabb(ba)^n \Leftrightarrow bb(ba)^n \Leftrightarrow (ba)^n bb$$

First, we write down an explicit rewrite path π_n for this equivalence.

Definition 4.2. For each $n \geq 0$, we inductively define the rewrite path π_n:

$$\begin{aligned} \pi_0 &:= \gamma_0 \\ \pi_{n+1} &:= aa(ba)^n b \cdot \beta^{-1} \triangleright \pi_n \cdot ba \triangleright (ba)^n b \cdot \beta \end{aligned}$$

We have $\mathsf{src}(\pi_n) = aa(ba)^n bb$, and $\mathsf{dst}(\pi_n) = (ba)^n bb$. An induction on n also shows that the path π_{n+1} is well-formed, in that the source of each step is in fact the destination of the previous step.

Now, let's take the three rules of R, and run Knuth-Bendix completion for "an infinitely long time." This process generates an infinite family of rules γ_n, where each rule is defined by the path π_n. More precisely:

Lemma 4.3. The monoid M admits an infinite complete presentation $\langle A \,|\, R^\infty\rangle$, where $R^\infty := \{\alpha, \beta\} \cup \{\gamma_n\}_{n \geq 0}$, and:

$$\begin{aligned} \alpha &:= (aaa, a) \\ \beta &:= (bba, abb) \\ \gamma_n &:= (aa(ba)^n bb, (ba)^n bb) \qquad \text{for all } n \geq 0 \end{aligned}$$

In the rest of this section, $\Rightarrow$ will denote the reduction relation generated by R^∞.

Proof. We must first show that $\langle A \,|\, R^\infty\rangle$ is a presentation of $\langle A \,|\, R\rangle$, and then establish that $\Rightarrow$ is terminating and confluent.

(Presentation.) Notice that $R = \{\alpha, \beta, \gamma_0\}$, so $R \subset R^\infty$, while $\mathsf{src}(\pi_n) = \mathsf{src}(\gamma_n)$ and $\mathsf{dst}(\pi_n) = \mathsf{dst}(\gamma_n)$, for all $n \geq 0$. It follows that R and R^∞ generate the same monoid congruence on A^*.

(Termination.) We use the shortlex order with $a < b$ to establish termination.

(1)	α vs. α:	aaa $\quad aaa$
(2)	α vs. α:	aaa $\quad aaa$
(3)	β vs. α:	bba $\quad aaa$
(4$_n$)	α vs. γ_n:	aaa $\quad aa(ba)^n bb$
(5$_n$)	α vs. γ_n:	aaa $\quad aa(ba)^n bb$
(6$_n$)	β vs. γ_n:	bba $\quad aa(ba)^n bb$
(7$_n$)	γ_n vs. β:	$aa(ba)^n bb$ $\qquad bba$
(8$_n$)	γ_n vs. β:	$aa(ba)^n bb$ $\qquad bba$

FIGURE 1. Overlapping left-hand sides in R^∞.

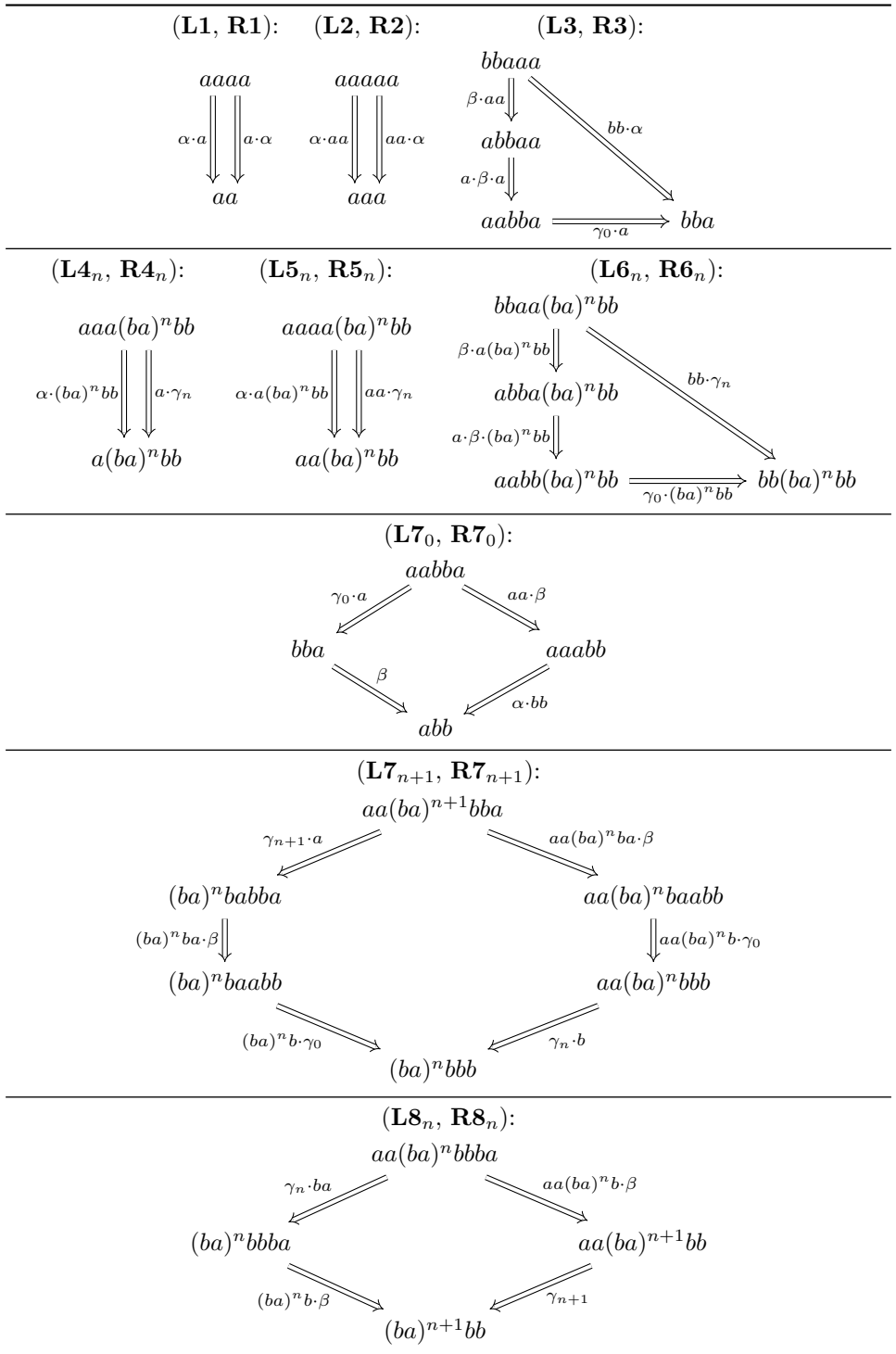


FIGURE 2. Resolution of critical pairs in $\Gamma(A \,|\, R^\infty)$.

(Confluence.) Figure 1 lists all overlapping rules in R^∞. We get the three critical pairs (**1**), (**2**), (**3**) by considering overlaps between α and β, together with five infinite families labeled ($\mathbf{4}_n$), ($\mathbf{5}_n$), ($\mathbf{6}_n$), ($\mathbf{7}_n$), and ($\mathbf{8}_n$), by considering overlaps involving the left-hand side of γ_n, and one of the other two rules.

Figure 2 shows that $\Rightarrow$ reduces both sides of each critical pair to the same word. We use the *leftmost* reduction strategy, where we always choose the positive rewrite step with the shortest left whisker. We label each pair of paths (**L1**, **R1**), (**L2**, **R2**), (**L3**, **R3**), ($\mathbf{L4}_n$, $\mathbf{R4}_n$), and so on. We observe that the rewrite paths resolving the critical pair ($\mathbf{7}_n$) take one form if $n = 0$, and another if $n > 0$, so we split the two cases into ($\mathbf{L7}_0$, $\mathbf{R7}_0$) and ($\mathbf{L7}_{n+1}$, $\mathbf{R7}_{n+1}$), to deal with them separately. □

In fact, $\langle A \,|\, R^\infty \rangle$ is a *regular* complete presentation of M, in that the set of left-hand sides of R^∞ is a *regular set*. We do not need to discuss regular sets in what follows, but one immediate corollary is that M has an easily decidable word problem, using *regular expressions*. Here is a Perl program to compute the R^∞-normal form of a list of words, given as command-line arguments:

```
foreach (@ARGV) {
    while (s/aaa/a/ || s/bba/abb/ || s/aa((ba)*bb)/\1/) {}
    print $_, "\n";
}
```

The `while` loop always terminates, and two words are equivalent if and only if they have the same R^∞-normal form.

The Degree Mapping

The reduction relation $\Rightarrow$ associated with $\langle A \,|\, R^\infty \rangle$ gives us a normal form for each equivalence class of words. We need to take a look at one specific property of these normal forms that we will need later.

Suppose that we have a pair of words with $x \Leftrightarrow y$, and x does not contain bb as a factor. Both sides of β and γ_0 contain bb, so the only rewrite step with source x must be an application of α. It follows that the destination of any step with source x also cannot have bb as a factor, so by induction we conclude the same about y.

Thus, the fun behaviors in M only manifest on equivalence classes of words that involve bb. Furthermore, when we compute the R^∞-normal form of a word, the distinct bb's accumulate at the end. For example, repeated application of β shows the following, with the right-hand side being irreducible:

$$(ba)^m bb (ba)^n bb \Rightarrow (ba)^{m+n} bbbb$$

Definition 4.4. If $x \in A^*$ is any word, the *degree* of x, denoted $\deg(x)$, is the largest integer $k \geq 0$ such that the R^∞-normal form of x ends with the suffix $(bb)^k$.

To continue the preceding example, we have:

$$\deg((ba)^m bb (ba)^n bb) = \deg((ba)^{m+n} bbbb) = 2$$

Remark 4.5. By definition, $x \Leftrightarrow y$ implies that $\deg(x) = \deg(y)$. However, "deg" is not a monoid homomorphism, because in general, $\deg(xy) \neq \deg(x) + \deg(y)$. One case where this fails in a strong way if we take $x := (ab)^k$ and $y := (ba)^k$. We have $\deg(x) = \deg(y) = 0$, while it is not hard to see that $xy \Leftrightarrow (bb)^k$, and so $\deg(xy) = k$.

Indeed, if $k = 0$, this is trivially true, and if $k > 0$, the claim follows by induction, because:

$$\begin{aligned} xy = (ab)^k(ba)^k &= (ab)^{k-1}abba(ba)^{k-1} \\ (\beta) \qquad &\Rightarrow (ab)^{k-1}aabb(ba)^{k-1} \\ (\gamma_0) \qquad &\Rightarrow (ab)^{k-1}bb(ba)^{k-1} \\ (\beta\text{, repeatedly}) \qquad &\Rightarrow (ab)^{k-1}(ba)^{k-1}bb \end{aligned}$$

Lemma 4.6. However, the following statements are true:

(1) If w has a suffix $(bb)^k$ for some $k \geq 0$, then $\deg(w) \geq k$.
(2) For all x, $y \in A^*$, we have $\deg(xy) \geq \deg(x) + \deg(y)$.

Proof. The first claim amounts to saying that $\Rightarrow$ only increases the number of bb's at the end of a word, so the R^∞-normal form has the longest suffix of bb's, among all words in the equivalence class of w. Thus, we assume that $w = w' \cdot (bb)^k$, and also that $w \Rightarrow z$; we must then show that $z = z' \cdot (bb)^k$, for some z'. By induction, it is sufficient to say this z is obtained from $w' \cdot (bb)^k$ by a single positive rewrite step. We consider each possibility in turn:

- For a step $x \cdot \alpha \cdot y$, we have $w = xaaay$ and $z = xay$, so $w' = xaaay'$ for some y', and $y = y' \cdot (bb)^k$, thus $z = xay' \cdot (bb)^k$.
- For a step $x \cdot \beta \cdot y$, we have $w = xbbay$ and $z = xabby$, and once again, $w' = xbbay'$ for some y', and $y = y' \cdot (bb)^k$, and thus $z = xabby' \cdot (bb)^k$.
- For a step $x \cdot \gamma_n \cdot y$, we have $w = xaa(ba)^n bby$, and $z = x(ba)^n bby$. We have $bby = y' \cdot (bb)^k$ for some y', so $z = x(ba)^n y' \cdot (bb)^k$.

The second claim is a consequence of the first. Suppose we are given a pair of words x, $y \in A^*$ such that $x \Rightarrow x'(bb)^m$ and $y \Rightarrow y'(bb)^n$ for some R^∞-irreducible words x', y', with m and n chosen to be maximal, so neither x' nor y' end with bb. By definition, we have $\deg(x) = m$ and $\deg(y) = n$. Also, $xy \Leftrightarrow x'y'(bb)^{m+n}$. While $x'y'$ need not be irreducible, by the first claim, we have $\deg(xy) \geq m+n$. □

A Homotopy Base

We can take each critical pair shown in Figure 1 and form a circuit in $C(A \,|\, R^\infty)$, by composing the right-hand side with the inverse of the left. By Lemma 3.11, this set of circuits is then an infinite homotopy base for $\langle A \,|\, R^\infty \rangle$. Our next goal is to transform this into a homotopy base for $\langle A \,|\, R \rangle$, by applying a mapping of rewrite paths. Recall the discussion around Lemma 3.15.

Definition 4.7. The isomorphism between the monoids presented by $\langle A \,|\, R^\infty \rangle$ and $\langle A \,|\, R \rangle$ gives us a mapping of rewrite paths $\Phi \colon P(A \,|\, R^\infty) \to P(A \,|\, R)$.

Under Φ, a rewrite step involving rule α, β, or γ_0 maps to the same step in $P(A \,|\, R)$, while the image of a step involving γ_n for $n > 0$ is the path π_n with appropriate whiskers:

$$\begin{aligned} \Phi(x \cdot \alpha \cdot y) &:= x \cdot \alpha \cdot y & \Phi(x \cdot \alpha^{-1} \cdot y) &:= x \cdot \alpha^{-1} \cdot y \\ \Phi(x \cdot \beta \cdot y) &:= x \cdot \beta \cdot y & \Phi(x \cdot \beta^{-1} \cdot y) &:= x \cdot \beta^{-1} \cdot y \\ \Phi(x \cdot \gamma_n \cdot y) &:= x \cdot \pi_n \cdot y & \Phi(x \cdot \gamma_n^{-1} \cdot y) &:= x \cdot \pi_n^{-1} \cdot y \end{aligned}$$

The mapping extends to arbitrary paths in the obvious way:

$$\Phi(s_1 \triangleright \cdots \triangleright s_n) = \Phi(s_1) \triangleright \cdots \triangleright \Phi(s_n)$$

$$
\begin{aligned}
\mathbf{C1} &= a \cdot \alpha \triangleright \alpha^{-1} \cdot a \\
\mathbf{C2} &= aa \cdot \alpha \triangleright \alpha^{-1} \cdot aa \\
\mathbf{C3} &= bb \cdot \alpha \triangleright \gamma_0^{-1} \cdot a \triangleright a \cdot \beta^{-1} \cdot a \triangleright \beta^{-1} \cdot aa \\
\mathbf{C4}_n &= a \cdot \pi_n \triangleright \alpha^{-1} \cdot (ba)^n bb \\
\mathbf{C5}_n &= aa \cdot \pi_n \triangleright \alpha^{-1} \cdot a(ba)^n bb \\
\mathbf{C6}_n &= bb \cdot \pi_n \triangleright \gamma_0^{-1} \cdot (ba)^n bb \triangleright a \cdot \beta^{-1} \cdot (ba)^n bb \triangleright \beta^{-1} \cdot a(ba)^n bb \\
\mathbf{C7}_0 &= aa \cdot \beta \triangleright \alpha \cdot bb \triangleright \beta^{-1} \triangleright \gamma_0^{-1} \cdot a \\
\mathbf{C7}_{n+1} &= aa(ba)^n ba \cdot \beta \triangleright aa(ba)^n b \cdot \gamma_0 \triangleright \pi_n \cdot b \\
&\qquad \triangleright (ba)^n b \cdot \gamma_0^{-1} \triangleright (ba)^n ba \cdot \beta^{-1} \triangleright \pi_{n+1}^{-1} \cdot a \\
\mathbf{C8}_n &= aa(ba)^n b \cdot \beta \triangleright \pi_{n+1} \triangleright (ba)^n b \cdot \beta^{-1} \triangleright \pi_n^{-1} \cdot ba
\end{aligned}
$$

TABLE 1. The infinite homotopy base $\mathcal{C}$.

The mapping Φ satisfies the conditions of the aforesaid lemma, so can apply Φ to each critical circuit shown in Figure 1, to get a homotopy base for $\langle A \,|\, R \rangle$.

Definition 4.8. Let $\mathcal{C}$ be the set of circuits obtained by applying Φ to each critical circuit of $\Gamma(A \,|\, R^\infty)$. We denote each circuit by **C1**, **C2**, **C3**, and so on:

$$\mathcal{C} := \{\mathbf{C1}, \mathbf{C2}, \mathbf{C3}, \mathbf{C7}_0\} \cup \{\mathbf{C4}_n, \mathbf{C5}_n, \mathbf{C6}_n, \mathbf{C7}_{n+1}, \mathbf{C8}_n\}_{n \geq 0}$$

Table 1 lists each rewrite path in $\mathcal{C}$.

Some of the circuits in $\mathcal{C}$ do not contribute anything to the generated homotopy relation. We're going to remove these redundant circuits from further consideration now, by showing that a smaller subset $\mathcal{B}$ of $\mathcal{C}$ is also a homotopy base for $\langle A \,|\, R \rangle$. This is going to involve some "trivial but technical" rewrite path algebra.

Lemma 4.9. The following subset $\mathcal{B} \subset \mathcal{C}$ is also a homotopy base for $\langle A \,|\, R \rangle$:

$$\mathcal{B} := \{\mathbf{C1}, \mathbf{C3}, \mathbf{C4}_0, \mathbf{C6}_0, \mathbf{C7}_0\} \cup \{\mathbf{C7}_{n+1}\}_{n \geq 0}$$

Proof. We claim that each circuit in $\mathcal{C} \setminus \mathcal{B}$ is null-homotopic to a combination of circuits from $\mathcal{B}$:

$$
\begin{aligned}
&(1) && \mathbf{C2} && \simeq a \cdot \mathbf{C1} \triangleright \mathbf{C1} \cdot a \\
&(2) && \mathbf{C4}_{n+1} && \simeq aaa(ba)^n b \cdot \beta^{-1} \triangleright \mathbf{C4}_n \cdot ba \triangleright aaa(ba)^n b \cdot \beta \\
&(3) && \mathbf{C5}_n && \simeq a \cdot \mathbf{C4}_n \triangleright \mathbf{C1} \cdot (ba)^n bb \\
&(4) && \mathbf{C6}_{n+1} && \simeq bbaa(ba)^n b \cdot \beta^{-1} \triangleright \mathbf{C6}_n \cdot ba \triangleright bbaa(ba)^n b \cdot \beta \\
&(5) && \mathbf{C8}_n && \simeq 1_{aa(ba)^n bbba}
\end{aligned}
$$

The conclusion will then follow by Lemma 3.8.

(1) We construct $\mathbf{C2}$ from two copies of $\mathbf{C1}$:

$$\begin{aligned}
\mathbf{C2} &= aa \cdot \alpha \triangleright \alpha^{-1} \cdot aa \\
\text{(inverses)} \qquad &\simeq aa \cdot \alpha \triangleright [a \cdot \alpha^{-1} \cdot a \triangleright a \cdot \alpha \cdot a] \triangleright \alpha^{-1} \cdot aa \\
&= [aa \cdot \alpha \triangleright a \cdot \alpha^{-1} \cdot a] \triangleright [a \cdot \alpha \cdot a \triangleright \alpha^{-1} \cdot aa] \\
&= a \cdot [a \cdot \alpha \triangleright \alpha^{-1} \cdot a] \triangleright [a \cdot \alpha \triangleright \alpha^{-1} \cdot a] \cdot a \\
&= a \cdot \mathbf{C1} \triangleright \mathbf{C1} \cdot a
\end{aligned}$$

(2) We construct $\mathbf{C4}_{n+1}$ from $\mathbf{C4}_n$, and thus $\mathbf{C4}_0$ by induction:

$$\begin{aligned}
\mathbf{C4}_{n+1} &= a \cdot \pi_{n+1} \triangleright \alpha^{-1} \cdot (ba)^{n+1}bb \\
&= aaa(ba)^n b \cdot \beta^{-1} \triangleright a \cdot \pi_n \cdot ba \triangleright [a(ba)^n b \cdot \beta \triangleright \alpha^{-1} \cdot (ba)^{n+1}bb] \\
\text{(interchange)} \qquad &\simeq aaa(ba)^n b \cdot \beta^{-1} \triangleright a \cdot \pi_n \cdot ba \triangleright [\alpha^{-1} \cdot (ba)^n bbba \triangleright aaa(ba)^n b \cdot \beta] \\
&= aaa(ba)^n b \cdot \beta^{-1} \triangleright [a \cdot \pi_n \cdot ba \triangleright \alpha^{-1} \cdot (ba)^n bbba] \triangleright aaa(ba)^n b \cdot \beta \\
&= aaa(ba)^n b \cdot \beta^{-1} \triangleright \mathbf{C4}_n \cdot ba \triangleright aaa(ba)^n b \cdot \beta
\end{aligned}$$

(3) We construct $\mathbf{C5}_n$ from $\mathbf{C4}_n$ (and thus $\mathbf{C4}_0$), together with $\mathbf{C1}$:

$$\begin{aligned}
\mathbf{C5}_n &= aa \cdot \pi_n \triangleright \alpha^{-1} \cdot a(ba)^n bb \\
\text{(inverses)} \qquad &\simeq aa \cdot \pi_n \triangleright [a \cdot \alpha^{-1} \cdot (ba)^n bb \triangleright a \cdot \alpha \cdot (ba)^n bb] \triangleright \alpha^{-1} \cdot a(ba)^n bb \\
&= [aa \cdot \pi_n \triangleright a \cdot \alpha^{-1} \cdot (ba)^n bb] \triangleright [a \cdot \alpha \cdot (ba)^n bb \triangleright \alpha^{-1} \cdot a(ba)^n bb] \\
&= a \cdot [a \cdot \pi_n \triangleright \alpha^{-1} \cdot (ba)^n bb] \triangleright [a \cdot \alpha \triangleright \alpha^{-1} \cdot a] \cdot (ba)^n bb \\
&= a \cdot \mathbf{C4}_n \triangleright \mathbf{C1} \cdot (ba)^n bb
\end{aligned}$$

(4) We construct $\mathbf{C6}_{n+1}$ from $\mathbf{C6}_n$, and thus $\mathbf{C6}_0$ by induction. First, we will look at this path:

$$\lambda_n := \gamma_0^{-1} \cdot (ba)^n bb \triangleright a \cdot \beta^{-1} \cdot (ba)^n bb \triangleright \beta^{-1} \cdot a(ba)^n bb$$

In fact, this is just the inverse of the left-hand side of the critical pair $(\mathbf{L6}_n, \mathbf{R6}_n)$, so $\mathbf{C6}_n = bb \cdot \pi_n \triangleright \lambda_n$. We observe that $\mathsf{src}(\lambda_n) = bb(ba)^n bb$ and $\mathsf{dst}(\lambda_n) = bbaa(ba)^n bb$. Furthermore, by interchange of disjoint rewrite steps:

$$bb(ba)^n b \cdot \beta \triangleright \lambda_{n+1} \simeq \lambda_n \cdot ba \triangleright bbaa(ba)^n b \cdot \beta$$

Now, we substitute π_{n+1} of Definition 4.2 into the definition of $\mathbf{C6}_{n+1}$ and make use of the above identity relating λ_{n+1} with λ_n:

$$\begin{aligned}
\mathbf{C6}_{n+1} &= [bb \cdot \pi_{n+1}] \triangleright \lambda_{n+1} \\
&= [bbaa(ba)^n b \cdot \beta^{-1} \triangleright bb \cdot \pi_n \cdot ba \triangleright bb(ba)^n b \cdot \beta] \triangleright \lambda_{n+1} \\
&= bbaa(ba)^n b \cdot \beta^{-1} \triangleright bb \cdot \pi_n \cdot ba \triangleright [bb(ba)^n b \cdot \beta \triangleright \lambda_{n+1}] \\
\text{(interchange)} \qquad &\simeq bbaa(ba)^n b \cdot \beta^{-1} \triangleright bb \cdot \pi_n \cdot ba \triangleright [\lambda_n \cdot ba \triangleright bbaa(ba)^n b \cdot \beta] \\
&= bbaa(ba)^n b \cdot \beta^{-1} \triangleright [bb \cdot \pi_n \cdot ba \triangleright \lambda_n \cdot ba] \triangleright bbaa(ba)^n b \cdot \beta \\
&= bbaa(ba)^n b \cdot \beta^{-1} \triangleright \mathbf{C6}_n \cdot ba \triangleright bbaa(ba)^n b \cdot \beta
\end{aligned}$$

(5) Finally, to see that the circuit $\mathbf{C8}_n$ is null-homotopic to the empty path at its basepoint, we substitute in π_{n+1} and cancel complementary rewrite steps:

$$\begin{aligned}
\mathbf{C8}_n &= aa(ba)^n b \cdot \beta \triangleright \pi_{n+1} \triangleright (ba)^n b \cdot \beta^{-1} \triangleright \pi_n^{-1} \cdot ba \\
&= aa(ba)^n b \cdot \beta \triangleright [aa(ba)^n b \cdot \beta^{-1} \triangleright \pi_n \cdot ba \triangleright (ba)^n b \cdot \beta] \\
&\quad \triangleright (ba)^n b \cdot \beta^{-1} \triangleright \pi_n^{-1} \cdot ba \\
&= [aa(ba)^n b \cdot \beta \triangleright aa(ba)^n b \cdot \beta^{-1}] \triangleright \pi_n \cdot ba \\
&\quad \triangleright [(ba)^n b \cdot \beta \triangleright (ba)^n b \cdot \beta^{-1}] \triangleright \pi_n^{-1} \cdot ba \\
\text{(inverses)} \qquad &\simeq \pi_n \cdot ba \triangleright \pi_n^{-1} \cdot ba \\
\text{(inverses)} \qquad &\simeq 1_{aa(ba)^n bbba}
\end{aligned}$$

Figure 3 gives a rewrite path for each circuit in $\mathcal{B}$, lists their basepoints, and illustrates each with a diagram. □

A Free Abelian Group

We now have our minimized homotopy base $\mathcal{B}$. However, $\mathcal{B}$ remains infinite. To show that no finite subset of $\mathcal{B}$ generates $\|$ as a homotopy relation, we define a "homomorphism" from rewrite paths into a certain free Abelian group.

We assume the reader has previously encountered free Abelian groups. We will follow the usual conventions and denote the commutative group operation by $+$, the inverse by $-$, and the identity by 0. The elements of a free Abelian group are finite formal sums of integer multiples of the generators.

Definition 4.10. Let G be the free Abelian group generated by all pairs $[\![u]\!] \otimes [\![v]\!]$ as $[\![u]\!]$ and $[\![v]\!]$ range over the elements of M (or in other words, the equivalence classes of $\Leftrightarrow$). That is, every element $z \in G$ has the form:

$$z = \sum_{i=1}^{n} c_i \cdot [\![u_i]\!] \otimes [\![v_i]\!]$$

for some $n \geq 0$, $c_i \in \mathbb{Z} \setminus \{0\}$, and u_i, $v_i \in A^*$.

The free monoid A^* acts on the left and right of G:

Definition 4.11. Suppose that u, v, $x \in A^*$. We define the left and right action of A^* on the generators of G as follows, and then extend to all of G by linearity:

$$\begin{aligned}
x \cdot ([\![u]\!] \otimes [\![v]\!]) &:= [\![xu]\!] \otimes [\![v]\!] \\
([\![u]\!] \otimes [\![v]\!]) \cdot x &:= [\![u]\!] \otimes [\![vx]\!]
\end{aligned}$$

It is important that the generators of G represent pairs of elements of M, and *not* words in A^*. For example, a typical element of G is $z := [\![aa]\!] \otimes [\![1]\!] - [\![1]\!] \otimes [\![1]\!]$. While certainly $z \neq 0$, the fact that $aaa \Leftrightarrow a$ means that $a \cdot z = 0$, because:

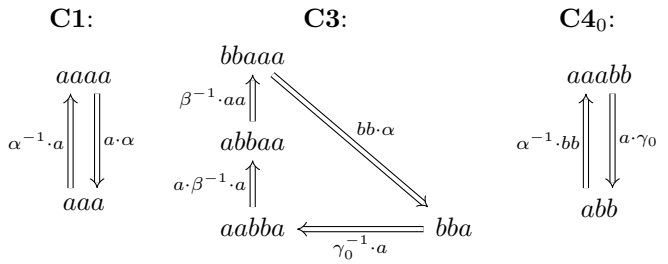
$$a \cdot z = [\![aaa]\!] \otimes [\![1]\!] - [\![a]\!] \otimes [\![1]\!] = 0$$

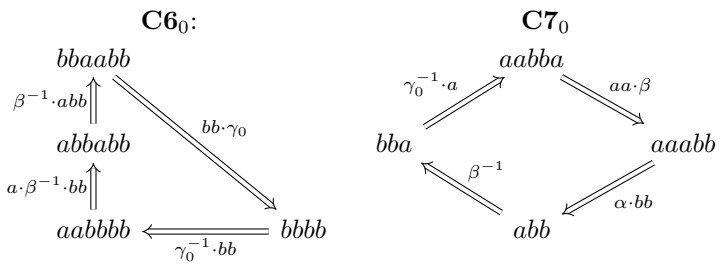
Every rewrite path in $P(A \,|\, R)$ maps to an element of G:

Definition 4.12. We define a mapping $\Theta\colon P(A \,|\, R) \to G$. Given a single rewrite step, we assign a value of 0 if the step applies rule α or β.

$$
\begin{aligned}
\mathbf{C1} &= a \cdot \alpha \triangleright \alpha^{-1} \cdot a \\
\mathbf{C3} &= bb \cdot \alpha \triangleright \gamma_0^{-1} \cdot a \triangleright a \cdot \beta^{-1} \cdot a \triangleright \beta^{-1} \cdot aa \\
\mathbf{C4}_0 &= a \cdot \gamma_0 \triangleright \alpha^{-1} \cdot bb \\
\mathbf{C6}_0 &= bb \cdot \gamma_0 \triangleright \gamma_0^{-1} \cdot bb \triangleright a \cdot \beta^{-1} \cdot bb \triangleright \beta^{-1} \cdot abb \\
\mathbf{C7}_0 &= aa \cdot \beta \triangleright \alpha \cdot bb \triangleright \beta^{-1} \triangleright \gamma_0^{-1} \cdot a \\
\mathbf{C7}_{n+1} &= aa(ba)^n ba \cdot \beta \triangleright aa(ba)^n b \cdot \gamma_0 \triangleright \pi_n \cdot b \\
&\qquad \triangleright (ba)^n b \cdot \gamma_0^{-1} \triangleright (ba)^n ba \cdot \beta^{-1} \triangleright \pi_{n+1}^{-1} \cdot a
\end{aligned}
$$

$$
\begin{aligned}
\mathsf{src}(\mathbf{C1}) &= aaaa & \mathsf{src}(\mathbf{C6}_0) &= bbaabb \\
\mathsf{src}(\mathbf{C3}) &= bbaaa & \mathsf{src}(\mathbf{C7}_0) &= aabba \\
\mathsf{src}(\mathbf{C4}_0) &= aaabb & \mathsf{src}(\mathbf{C7}_{n+1}) &= aa(ba)^{n+1}bba
\end{aligned}
$$





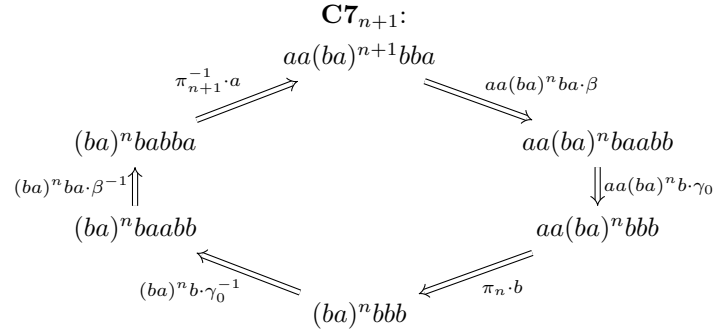


FIGURE 3. The minimized homotopy base $\mathcal{B} \subset \mathcal{C}$.

For a step that applies γ_0, we form an element of G by pairing the step's left and right whiskers, and negate the result if the step is negative:

$$\begin{aligned} \Theta(x \cdot \alpha \cdot y) &:= 0 & \Theta(x \cdot \alpha^{-1} \cdot y) &:= 0 \\ \Theta(x \cdot \beta \cdot y) &:= 0 & \Theta(x \cdot \beta^{-1} \cdot y) &:= 0 \\ \Theta(x \cdot \gamma_0 \cdot y) &:= [\![x]\!] \otimes [\![y]\!] & \Theta(x \cdot \gamma_0^{-1} \cdot y) &:= -[\![x]\!] \otimes [\![y]\!] \end{aligned}$$

Finally, we extend Θ to arbitary paths $p := s_1 \triangleright \cdots \triangleright s_n$, by linearity:

$$\Theta(p) := \Theta(s_1) + \cdots + \Theta(s_n)$$

The mapping Θ "respects" the algebra of rewrite paths. From the definition, we see that for $p, q \in P(A \mid R)$ with $\mathsf{dst}(p) = \mathsf{src}(q)$, and arbitrary words $x, y \in A^*$:

$$\begin{aligned} \Theta(1_x) &= 0 \\ \Theta(p \triangleright q) &= \Theta(p) + \Theta(q) \\ \Theta(p^{-1}) &= -\Theta(p) \\ \Theta(x \cdot p \cdot y) &= x \cdot \Theta(p) \cdot y \end{aligned}$$

A further consequence of the above is that Θ is blind to null homotopy:

Lemma 4.13. Suppose that $p, q \in P(A \mid R)$ satisfy $p \simeq q$. Then $\Theta(p) = \Theta(q)$.

Proof. By Lemma 3.2, there exist a sequence of paths $p_1, \ldots, p_n$ such that $p = p_1$, $q = p_n$, and each p_{i+1} is obtained from p_i by inserting or removing complementary rewrite steps, or by interchanging disjoint rewrite steps. We will see in each case $\Theta(p_i) = \Theta(p_{i+1})$, and so our conclusion will follow by induction on n.

If p_{i+1} is obtained from p_i by inserting or removing complementary rewrite steps $s \triangleright s^{-1}$, we certainly have $\Theta(p_{i+1}) = \Theta(p_i)$, because $\Theta(s \triangleright s^{-1}) = 0$ for any rewrite step s.

Now, if p_{i+1} is obtained from p_i by interchanging disjoint rewrite steps, then we're replacing $s_1 \cdot \mathsf{src}(s_2) \triangleright \mathsf{dst}(s_1) \cdot s_2$ with $\mathsf{src}(s_1) \cdot s_2 \triangleright s_1 \cdot \mathsf{dst}(s_2)$, or vice versa. If we apply Θ to both sides, we get:

$$\begin{aligned} \Theta(s_1 \cdot \mathsf{src}(s_2) \triangleright \mathsf{dst}(s_1) \cdot s_2) &= \Theta(s_1) \cdot \mathsf{src}(s_2) + \mathsf{dst}(s_1) \cdot \Theta(s_2) \\ \Theta(\mathsf{src}(s_1) \cdot s_2 \triangleright s_1 \cdot \mathsf{dst}(s_2)) &= \Theta(s_1) \cdot \mathsf{dst}(s_2) + \mathsf{src}(s_1) \cdot \Theta(s_2) \end{aligned}$$

But $\mathsf{src}(s_1) \Leftrightarrow \mathsf{dst}(s_1)$, and $\mathsf{src}(s_2) \Leftrightarrow \mathsf{dst}(s_2)$, so $\Theta(p_i) = \Theta(p_{i+1})$. □

Table 2 lists the image under Θ of each circuit in $\mathcal{B}$ from Figure 3. Recall that $\mathbf{C7}_{n+1}$ is expressed in terms of π_n, so to get $\Theta(\mathbf{C7}_{n+1})$, we first need:

Lemma 4.14. For all $n \geq 0$, we have $\Theta(\pi_n) = [\![1]\!] \otimes [\![(ba)^n]\!]$.

Proof. In the base case, $\Theta(\pi_0) = \Theta(\gamma_0) = [\![1]\!] \otimes [\![1]\!]$. Otherwise, we assume that $\Theta(\pi_n) = [\![1]\!] \otimes [\![(ba)^n]\!]$. We show that the same statement holds for $n+1$, by substituting in Definition 4.2 for $n+1$:

$$\begin{aligned} \Theta(\pi_{n+1}) &= \Theta(aa(ba)^n b \cdot \beta^{-1}) + \Theta(\pi_n \cdot ba) + \Theta((ba)^n b \cdot \beta) \\ &= \Theta(\pi_n) \cdot ba \\ &= ([\![1]\!] \otimes [\![(ba)^n]\!]) \cdot ba \\ &= [\![1]\!] \otimes [\![(ba)^{n+1}]\!] \end{aligned}$$

The conclusion follows for all $n \geq 0$ by induction. □

$$\begin{aligned}
\Theta(\mathbf{C1}) &= 0\\
\Theta(\mathbf{C3}) &= -[\![1]\!] \otimes [\![a]\!]\\
\Theta(\mathbf{C4}_0) &= [\![a]\!] \otimes [\![1]\!]\\
\Theta(\mathbf{C7}_0) &= -[\![1]\!] \otimes [\![a]\!]\\
\Theta(\mathbf{C6}_0) &= [\![bb]\!] \otimes [\![1]\!] - [\![1]\!] \otimes [\![bb]\!]\\
\Theta(\mathbf{C7}_{n+1}) &= ([\![aa(ba)^n b]\!] \otimes [\![1]\!] - [\![(ba)^n b]\!] \otimes [\![1]\!])\\
&\quad + ([\![1]\!] \otimes [\![(ba)^n b]\!] - [\![1]\!] \otimes [\![(ba)^n baa]\!])
\end{aligned}$$

TABLE 2. Applying Θ to each element of our homotopy base $\mathcal{B}$.

MINIMALITY

Finally, we reach the crucial point in the proof of Theorem 1. We bring together the degree map from Definition 4.4, the map Θ from Definition 4.12, and the homotopy base $\mathcal{B}$ from Lemma 4.9, to show that no finite subset of $\mathcal{B}$ is a homotopy base.

Proof of Theorem 1. Suppose that, for the sake of contradiction, there is a finite subset $\mathcal{B}' \subset \mathcal{B}$ such that $\mathcal{B}'$ is a homotopy base for $\langle A \mid R \rangle$.

Since $\mathcal{B}'$ is finite, there exists a $k \geq 0$ such that $\mathbf{C7}_{k+1} \notin \mathcal{B}'$. Since $\mathcal{B}'$ is a homotopy base, by Lemma 3.8, we can construct $\mathbf{C7}_{k+1}$ from the circuits in $\mathcal{B}'$, meaning there exist words $x_i, y_i \in A^*$, paths $p_i \in P(A \mid R)$, circuits $q_i \in \mathcal{B}'$, and integers $e_i \in \{-1, 1\}$, such that:

$$\mathbf{C7}_{k+1} \simeq p_1 \triangleright (x_1 \cdot q_1^{e_1} \cdot y_1) \triangleright p_1^{-1} \triangleright \cdots \triangleright p_n \triangleright (x_n \cdot q_n^{e_n} \cdot y_n) \triangleright p_n^{-1} \tag{1}$$

By Lemma 4.13, we can apply Θ to both sides of equation (1). On the right-hand side, each term $\Theta(p_i)$ cancels with the corresponding $\Theta(p_i^{-1})$, and what remains is a summand for each q_i, whiskered by the x_i and y_i:

$$\Theta(\mathbf{C7}_{k+1}) = \sum_{i=1}^{n} \Theta(x_i \cdot q_i^{e_i} \cdot y_i) \tag{2}$$

To get a contradiction from this, we prove a series of claims. The first claim is that the left-hand side of (2) is *not* an element of a certain subgroup $H \subset G$.

Definition 4.15. Let $H \subset G$ denote the subgroup of G generated by elements of the form $[\![xa]\!] \otimes [\![y]\!]$ and $[\![x]\!] \otimes [\![ay]\!]$, as x, y range over all equivalence classes where $\deg(x) = 0$ and $\deg(y) = 0$. Note that $\deg(x) = \deg(y) = 0$ is equivalent to saying that $\deg(xa) = \deg(ay) = 0$, and since every subgroup of a free Abelian group is free Abelian, it also the case that H is free Abelian.

Claim. For all $k \geq 0$, we have $\Theta(\mathbf{C7}_{k+1}) \notin H$.

Proof. Suppose that to the contrary, we can express $\Theta(\mathbf{C7}_{k+1})$ in terms of the generators of H, so there are integers c_i, c'_j, and degree zero words x_i, y_i, x'_j, y'_j, such that:

$$\Theta(\mathbf{C7}_{k+1}) = \sum_{i=1}^{m} c_i \cdot [\![x_i a]\!] \otimes [\![y_i]\!] + \sum_{j=1}^{n} c'_j \cdot [\![x'_j]\!] \otimes [\![a y'_j]\!]$$

We quickly see this is impossible. We refer to Table 2, and consider any summand of $\Theta(\mathbf{C7}_{k+1})$, such as $[\![(ba)^k b]\!] \otimes [\![1]\!]$. There must exist an $1 \leq i \leq m$ or $1 \leq j \leq n$, such that:

$$[\![(ba)^k b]\!] \otimes [\![1]\!] = [\![x_i a]\!] \otimes [\![y_i]\!] \qquad \text{or} \qquad [\![(ba)^k b]\!] \otimes [\![1]\!] = [\![x_i']\!] \otimes [\![a y_i']\!]$$

If the first case were to hold, we would have $(ba)^k b \Leftrightarrow x_i a$. However, this is an equivalence between degree zero words, so the only rule we can apply is α, and a rewrite step that applies α cannot flip the last letter of a word from b to a. In the second case, we get $1 \Leftrightarrow a y_i'$. Here again we have a problem, because the equivalence class of 1 contains no other words but the empty word itself. In any case, $\Theta(\mathbf{C7}_{n+1}) \notin H$. □

The remaining claims show that every possible summand of the right-hand side of equation (2) *is* an element of H, which is our contradiction.

These claims all rely on the fact that every word visited by the circuit on the right-hand side of (1) must be in the equivalence class of $\mathsf{src}(\mathbf{C7}_{k+1}) = aa(ba)^{k+1}bba$. In particular, this constrains the possibilities for each pair of whiskers x_i and y_i, because for all i, we must have:

$$\mathsf{src}(\mathbf{C7}_{k+1}) \Leftrightarrow x_i \cdot \mathsf{src}(q_i^{e_i}) \cdot y_i$$

Furthermore, $\deg(\mathsf{src}(\mathbf{C7}_{k+1})) = 1$, so for all i in equation (2), we also have:

$$\deg(x_i \cdot \mathsf{src}(q_i) \cdot y_i) = 1$$

Lemma 4.6 then implies that $\deg(\mathsf{src}(q_i)) \leq 1$. An immediate consequence:

Claim. The circuit $\mathbf{C6}_0$ does not appear on the right-hand side of equation (2).

Proof. Since $bbaabb \Rightarrow bbbb$, we have $\deg(\mathsf{src}(\mathbf{C6}_0)) = \deg(bbbb) = 2$. □

Now, we consider each remaining possibility for q_i.

Claim. If $q_i = \mathbf{C1}$ for some i in equation (2), then $\Theta(x_i \cdot q_i \cdot y_i) = 0$.

Proof. Recall from Table 2 that $\Theta(\mathbf{C1}) = 0$, because $\mathbf{C1}$ does not involve γ_0. □

Claim. If $q_i \in \{\mathbf{C3}, \mathbf{C4}_0, \mathbf{C7}_0\}$ for some i in equation (2), then $\Theta(x_i \cdot q_i \cdot y_i) \in H$.

Proof. Let $x := x_i$ and $y := y_i$ to avoid clutter. If q_i is as above, then inspection of Figure 3 shows $\deg(\mathsf{src}(q_i)) = 1$. Together with $\deg(x_i \cdot \mathsf{src}(q_i) \cdot y_i) = 1$, this forces $\deg(x_i) = \deg(y_i) = 0$. Now, from Table 2:

$$\begin{aligned}
\Theta(x \cdot \mathbf{C3} \cdot y) &= -[\![x]\!] \otimes [\![ay]\!] \in H \\
\Theta(x \cdot \mathbf{C4}_0 \cdot y) &= [\![xa]\!] \otimes [\![y]\!] \in H \\
\Theta(x \cdot \mathbf{C7}_0 \cdot y) &= -[\![x]\!] \otimes [\![ay]\!] \in H
\end{aligned}$$

The fact that these are all in H follows directly from Definition 4.15. □

Hence the q_i appearing on the right-hand side of equation (2) cannot be solely drawn from the set $\{\mathbf{C1}, \mathbf{C3}, \mathbf{C4}_0, \mathbf{C7}_0\}$, because then the entire right-hand side would be an element of H, which would contradict $\Theta(\mathbf{C7}_{k+1}) \notin H$.

We still haven't ruled out the possibility that equation (2) has at least one summand $q_i = \mathbf{C7}_{m+1}$ for some other $m \geq 0$. Perhaps $\mathcal{B}'$ simply contains some finite number of circuits of this form? However, this is again insufficient. When $m \neq k$, the whiskers x_i and y_i cannot both be empty, and a degree argument slightly more elaborate than the previous claim will again imply that $\Theta(x_i \cdot q_i \cdot y_i) \in H$.

Claim. If $q_i = \mathbf{C7}_{m+1}$ for some $m \geq 0$, then $\Theta(x_i \cdot q_i \cdot y_i) \in H$.

Proof. Let $x := x_i$, and $y := y_i$ to avoid clutter.

We know that $\mathbf{C7}_{k+1} \notin \mathcal{B}'$, so $m \neq k$. Therefore, $\mathsf{src}(\mathbf{C7}_{k+1}) \not\Leftrightarrow \mathsf{src}(\mathbf{C7}_{m+1})$. However, we still have $\mathsf{src}(\mathbf{C7}_{k+1}) \Leftrightarrow x \cdot \mathsf{src}(\mathbf{C7}_{m+1}) \cdot y$, which forces $|x| + |y| > 0$. As before, $\deg(\mathsf{src}(\mathbf{C7}_{m+1})) = 1$ also forces $\deg(x) = \deg(y) = 0$.

If $|x| > 0$, then x cannot end with b, because if $x = x'b$ for some $x' \in A^*$, it would imply that $\deg(x'b \cdot \mathsf{src}(\mathbf{C7}_{m+1}) \cdot y) \geq 2$, since $\deg(b \cdot \mathsf{src}(\mathbf{C7}_{m+1})) = 2$:

$$\begin{aligned} b \cdot \mathsf{src}(\mathbf{C7}_{m+1}) &= baa(ba)^{m+1}bba \\ (\gamma_{m+1}) \qquad &\Rightarrow bb(ab)^m abba \\ (\beta) \qquad &\Rightarrow bb(ab)^m aabb \\ (\gamma_0) \qquad &\Rightarrow bb(ab)^m bb \\ (\beta\text{, repeatedly}) \qquad &\Rightarrow (ab)^m bbbb \end{aligned}$$

Similarly, if $|y| > 0$, then y cannot start with b, because if $y = by'$ for some $y' \in A^*$, it would imply that $\deg(x \cdot \mathsf{src}(\mathbf{C7}_{m+1}) \cdot by') \geq 2$, since $\deg(\mathsf{src}(\mathbf{C7}_{m+1}) \cdot b) = 2$:

$$\begin{aligned} \mathsf{src}(\mathbf{C7}_{m+1}) \cdot b &= aa(ba)^{m+1}bbab \\ (\beta) \qquad &\Rightarrow aa(ba)^m baabbb \\ (\gamma_0) \qquad &\Rightarrow aa(ba)^m bbbb \\ (\gamma_m) \qquad &\Rightarrow (ba)^m bbbb \end{aligned}$$

Three possibilities remain:

(1) If $x = x'a$ and $y = 1$, we see that two of the terms in $\Theta(\mathbf{C7}_{m+1})$ cancel out, because $aaa \Leftrightarrow a$, and the remaining two terms are elements of H:

$$\begin{aligned} \Theta(x \cdot \mathbf{C7}_{m+1} \cdot y) &= [\![x'aaa(ba)^m b]\!] \otimes [\![1]\!] + [\![x'a]\!] \otimes [\![(ba)^m b]\!] \\ &\quad - [\![x'a(ba)^m b]\!] \otimes [\![1]\!] - [\![x'a]\!] \otimes [\![(ba)^m baa]\!] \\ &= [\![x'a]\!] \otimes [\![(ba)^m b]\!] - [\![x'a]\!] \otimes [\![(ba)^m baa]\!] \in H \end{aligned}$$

(2) If $x = 1$ and $y = ay'$ instead, the other two terms cancel out, and we reach the same conclusion:

$$\begin{aligned} \Theta(x \cdot \mathbf{C7}_{m+1} \cdot y) &= [\![aa(ba)^m b]\!] \otimes [\![ay']\!] + [\![1]\!] \otimes [\![(ba)^m bay']\!] \\ &\quad - [\![(ba)^m b]\!] \otimes [\![ay']\!] - [\![1]\!] \otimes [\![(ba)^m baaay']\!] \\ &= [\![aa(ba)^m b]\!] \otimes [\![ay']\!] - [\![(ba)^m b]\!] \otimes [\![ay']\!] \in H \end{aligned}$$

(3) Finally, if $x = x'a$ and $y = ay'$, everything cancels out:

$$\begin{aligned} \Theta(x \cdot \mathbf{C7}_{m+1} \cdot y) &= [\![x'aaa(ba)^m b]\!] \otimes [\![ay']\!] + [\![x'a]\!] \otimes [\![(ba)^m bay']\!] \\ &\quad - [\![x'a(ba)^m b]\!] \otimes [\![ay']\!] - [\![x'a]\!] \otimes [\![(ba)^m baaay']\!] \\ &= 0 \in H \end{aligned}$$

In any case, $\Theta(x \cdot \mathbf{C7}_{m+1} \cdot y) \in H$. □

Thus, claims two through five show that the right-hand side of equation (2) must be in H, which contradicts the first claim. We conclude that if $\mathcal{B}' \subset \mathcal{B}$, and $\mathbf{C7}_{k+1} \notin \mathcal{B}'$ for at least one $k \geq 0$, then $\mathcal{B}'$ is not a homotopy base for $\langle A \mid R \rangle$. □

Total after symmetry:		19,575
Solved by shortlex $a < b$:		19,348
Trivial with 1 element:	1,188	
Finite with > 1 element:	11,723	
Infinite:	6,437	
Solved by other reduction orders:		34
Solved by adding a generator:		190
Remaining unsolved:		3

TABLE 3. Two-generator, two-relation monoids of length ≤ 10.

5. A Monoid Census

This section describes an investigation, loosely inspired by Bogdan Grechuk's wonderful book on the topic of *Diophantine equations* [21]. Like the word problem, no general approach can solve all Diophantine equations, because the general case is undecidable. Grechuk defines an ordering of all such equations, and then applies various techniques to solve as many as possible, in order of increasing size.

Similarly, the non-FDT monoid $\langle a, b \,|\, aaa = a, abba = bb\rangle$ was "discovered" by enumerating all monoid presentations with two generators and two relations up to length 10, and applying the Knuth-Bendix algorithm to each one in turn. The code is in Swift compiler repository, in the form of a performance micro-benchmark:

`https://github.com/swiftlang/swift/tree/main/benchmark/multi-source/Monoids`

The program finds a complete presentation for all but three instances in the enumeration, using 3 seconds of real time and 38 seconds of CPU time on an Apple MacBook Pro with 14-core M4 chip, which suggests the feasability of exploring larger enumerations. The program can also be compiled as a standalone binary, which prints results to standard output. The output of a run can be found here:

`https://factorcode.org/slava/monoids2210.txt`

Table 3 shows a summary of the results, which will be explained below.

The first step is to filter out certain instances. While no attempt was made to classify the monoids up to isomorphism, some obviously isomorphic instances can be eliminated early on to avoid unnecessary work. Let's say that two monoid presentations are equivalent if we can get one from the other by some combination of the following:

(1) Swapping the two sides of a defining relation.
(2) Swapping the two defining relations.
(3) Replacing a with b, and vice versa, within each defining relation.

If one instance in such an equivalence class admits a finite complete presentation, they all do, so the filtering step selects one representative.

We also skip instances where a defining relation is the tautology $u = u$, or one of the form $u = v$ where both $|u| \leq 1$ and $|v| \leq 1$. (This eliminates one instance in particular, $\langle a, b \,|\, abbaab = ba, a = a\rangle$, which reduces to a hard one-relation monoid the author has yet to figure out. However, every one-relation monoid has FDT [16], and the immediate goal here was to look for something not FDT.)

At this point, we are ready to start solving in earnest. We print the total number of instances that must be solved after accounting for symmetry:

```
# Remaining 19575
```

We then print some column headings:

```
# n:  presentation:  cardinality:  complete:  strategy:
```

The main loop kicks off Knuth-Bendix completion on each remaining instance, with parameters dictated by the current strategy. We use Swift's `Task` API to parallelize the work, by running 32 tasks at a time. Each task attempts completion on a single instance. If we get a complete presentation within a fixed number of iterations, we retire the instance and proclaim that it has been solved, by printing out a row of output for each of the column headings above. These consist of:

(1) The instance number.
(2) The defining relations of the presentation.
(3) The cardinality of the presented monoid, or `infinite` if it is infinite.
(4) A finite complete presentation for the presented monoid.
(5) The generators added to get a finite complete presentation, if any.

The first strategy attempts Knuth-Bendix completion with the original alphabet $\{a, b\}$, and the shortlex order $a < b$. This solves 19,348 instances—almost all.

Example 1. Let's look at the first four instances. The first two present $\mathbb{Z}_2$,

```
1  aa=1,ab=1  finite:2  fcrs:b=a,aa=1
2  aa=1,ba=1  finite:2  fcrs:b=a,aa=1
```

then we have their free product, $\mathbb{Z}_2 * \mathbb{Z}_2$,

```
3  aa=1,bb=1  infinite  fcrs:aa=1,bb=1
```

which is followed by the free Abelian group $\mathbb{Z}$:

```
4  ab=1,ba=1  infinite  fcrs:ab=1,ba=1
```

Finite Monoids

Notice how the first two monoids in our enumeration are *finite*. Every finite monoid has a finite complete presentation [22]. From a finite complete presentation, we can decide if the presented monoid is finite, and compute its cardinality, as follows:

(1) From Section 2, a word is irreducible with respect to a presentation $\langle A \mid R \rangle$ if it does not contain the left-hand side of a rule from R as a factor.
(2) If the set R is finite, the set of irreducible words is a regular set, so it can be recognized by a finite state automaton (for definitions, see [23]).
(3) We construct this automaton using the algorithm from Lemma 2.1.3 of [18].
(4) We first check if this automaton contains a cycle, in which case the set of irreducible words is infinite.
(5) Otherwise, we count the number of words accepted by this automaton. This gives us the number of elements in the presented monoid.

In our enumeration, there are 12911 presentations of finite monoids. All can be solved by the first pass.

This includes 1188 presentations of the trivial monoid. The finite complete presentation of the trivial monoid $\langle a, b \mid a = 1, b = 1 \rangle$ does not itself appear in our enumeration, because as we mentioned, we discard instances with defining relations where both sides have length ≤ 1. However, we don't have to look far to find other presentations of the trivial monoid.

1: 1188, 2: 2059, 3: 1233, 4: 1644, 5: 1019, 6: 1630, 7: 686, 8: 884, 9: 493, 10: 615, 11: 191, 12: 317, 13: 79, 14: 134, 15: 62, 16: 158, 17: 16, 18: 60, 19: 2, 20: 52, 21: 38, 22: 12, 24: 31, 25: 5, 26: 11, 27: 42, 28: 10, 30: 52, 32: 5, 34: 8, 36: 17, 39: 4, 42: 4, 44: 8, 48: 12, 50: 12, 52: 4, 55: 4, 56: 8, 60: 18, 64: 14, 81: 6, 84: 22, 96: 1, 100: 4, 105: 2, 120: 7, 129: 2, 147: 6, 160: 2, 165: 2, 192: 2, 195: 2, 320: 2, 324: 2, 339: 4, 605: 2, 1083: 2.

TABLE 4. Number of finite monoids of each order.

Example 2. The first presentation of the trivial monoid is number 5:

```
5  aa=a,ab=1  finite:1  fcrs:a=1,b=1
```

From this presentation $\langle a, b \,|\, aa = a, ab = 1\rangle$, it easy to see that $a \Leftrightarrow aab \Leftrightarrow ab \Leftrightarrow 1$ and therefore $b \Leftrightarrow ab \Leftrightarrow 1$ also.

Example 3. A more interesting example is instance number 8216:

```
8216  abab=b,bbaaa=1  finite:1  fcrs:a=1,b=1
```

Here is the shortest proof the author could find that $\langle a, b \,|\, abab = b, bbaaa = 1\rangle$ presents the trivial monoid. First, we have:

$$a \Leftrightarrow abbaaa \Leftrightarrow abbaabbaaaa \Leftrightarrow abbaababababaaaa \Leftrightarrow abbababaaaa$$
$$\Leftrightarrow abbbaaaa \Leftrightarrow aba \Leftrightarrow ababbaaa \Leftrightarrow bbaaa \Leftrightarrow 1$$

Thus, we can remove the letter a from the presentation, leaving us with $bb \Leftrightarrow b$ and $bb \Leftrightarrow 1$, from which it follows that $b \Leftrightarrow 1$ also.

Table 4 shows a histogram with the number of presentations of finite monoids of each size. Table 5 lists the largest handful of instances.

Example 4. The finite monoids with 324 elements have a simple description. We will look at $\langle a, b \,|\, aaab = ba, bbbb = 1\rangle$; the other one is anti-isomorphic. We have $a \Leftrightarrow a^{81}$, because:

$$a \Leftrightarrow b^4a \Leftrightarrow b^3a^3b \Leftrightarrow b^2a^9b^2 \Leftrightarrow ba^{27}b^3 \Leftrightarrow a^{81}b^4 \Leftrightarrow a^{81}$$

Therefore, given any word, we can collect all the b's at the end. This can blow up the number of a's, but then we repeatedly replace a^{81} with a, and b^4 with 1, until we get a word of the form $a^m b^n$, where $0 \leq m < 81$ and $0 \leq n < 4$.

Indeed, $81 \cdot 4 = 324$ is the order of the presented monoid.

The maximum is achieved by two instances with a whopping 1083 elements:

Question 1. Is there a simple way to describe $\langle a, b \,|\, aaa = 1, bbbb = aba\rangle$ and $\langle a, b \,|\, aaa = 1, abbbba = b\rangle$, with 1083 elements? Are they isomorphic?

OTHER REDUCTION ORDERS

The first pass fails with 227 instances, meaning that Knuth-Bendix hits the iteration limit without success. As mentioned in Section 1, it is well-known that successful completion may depend on a choice of reduction order [7]. We can observe this in our enumeration.

320 elements:	$\langle a, b \mid aaaa = 1, babbb = a\rangle$	$\langle a, b \mid aaaa = 1, bbbab = a\rangle$
324 elements:	$\langle a, b \mid aaab = ba, bbbb = 1\rangle$	$\langle a, b \mid baaa = ab, bbbb = 1\rangle$
339 elements:	$\langle a, b \mid aaa = 1, ababba = b\rangle$ $\langle a, b \mid aaa = 1, babb = aba\rangle$	$\langle a, b \mid aaa = 1, abbaba = b\rangle$ $\langle a, b \mid aaa = 1, bbab = aba\rangle$
605 elements:	$\langle a, b \mid abba = b, aaaaa = 1\rangle$	$\langle a, b \mid aba = bb, aaaaa = 1\rangle$
1083 elements:	$\langle a, b \mid aaa = 1, bbbb = aba\rangle$	$\langle a, b \mid aaa = 1, abbbba = b\rangle$

TABLE 5. The largest finite monoids in our enumeration.

Example 5. Here is instance number 7499:

$$\langle a, b \mid aba = aab, bab = b\rangle$$

The first pass fails to solve this instance because Knuth-Bendix completion does not terminate with the shortlex order $a < b$, instead enumerating the rules of this infinite complete presentation:

$$\begin{aligned} aba &\Rightarrow aab \\ aabb &\Rightarrow ab \\ ba^n ab &\Rightarrow ba^n \qquad \text{for all } n \geq 0 \end{aligned}$$

However, in this case, simply orienting the initial rules with the shortlex order $b < a$ is sufficient to obtain a finite complete presentation. Knuth-Bendix does not have to add any more rules, because all critical pairs resolve immediately:

$$\begin{aligned} bab &\Rightarrow b \\ aab &\Rightarrow aba \end{aligned}$$

Thus, we perform three additional passes, to attempt the following additional reduction orders on all remaining instances:

- Shortlex $b < a$.
- Recursive path $a \wr b$.
- Recursive path $b \wr a$. (The recursive path order is described in [24].)

This solves 34 more instances, but 193 still remain.

ADDING A GENERATOR

The remaining instances do not admit a finite complete presentation over $\{a, b\}$ compatible with any of our reduction orders, but most still admit finite complete presentations over a different alphabet. This phenomenon was first noted in [6].

Example 6. Here is instance number 10717:

$$\langle a, b \mid abab = 1, abbba = b\rangle$$

Knuth-Bendix over $\{a, b\}$ fails with each one of our reduction orders. For example, with shortlex $a < b$, we get this infinite complete presentation:

$$\begin{aligned} bba &\Rightarrow abb \\ abab &\Rightarrow 1 \\ baba &\Rightarrow 1 \\ aba^{n+1}bb &\Rightarrow ba^n \qquad \text{for all } n \geq 1 \\ ba^n ba &\Rightarrow aba^n b \quad \text{for all } n \geq 2 \end{aligned}$$

On the other hand, if we add a generator c and a defining relation (aba, c), we get a different presentation of the same monoid:

$$\langle a, b, c \,|\, abab = 1, abbba = b, aba = c \rangle$$

Knuth-Bendix successfully completes with the recursive path order $b \wr c \wr a$, and we get this finite complete presentation over $\{a, b, c\}$:

$$\begin{aligned} cb &\Rightarrow 1 \\ bc &\Rightarrow 1 \\ bba &\Rightarrow abb \\ ca &\Rightarrow bacc \\ aba &\Rightarrow c \end{aligned}$$

Notice in the preceding example, the failed run over the alphabet $\{a, b\}$ involves rules with many occurrences of the factor aba, and then introducing (aba, c) gives us a finite complete presentation. This suggests the following simple heuristic for guessing the extra generator, inspired by [25].

First, we perform two rounds of Knuth-Bendix completion, and stop. The result is not confluent, but we can collect all factors of all the words appearing in the rules added so far, and order them by frequency. In the current enumeration, it suffices to take the highest and second-highest frequency factor, of each length from 2 to 6. (The author has also experimented with heuristics for adding multiple generators, but this is not required for any instances in this enumeration.)

For each remaining instance, and for each such factor $u \in \{a, b\}^*$ in the factor set of this instance, we add a new letter c and rule (u, c) to the original presentation, to get a different presentation of the same monoid. We attempt completion on this isomorphic presentation, this time with every possible shortlex and recursive path order on the extended alphabet $\{a, b, c\}$. This solves 190 additional instances.

Unsolved Instances

Only 3 instances remain, all with length 10. We print them before exiting:

```
# Remaining 3
7415   bab=aaa,bbbb=1  hard
10397  aaaa=1,abbba=b  hard
11931  aaa=a,abba=bb   hard
```

The third instance is our friend $\langle a, b \,|\, aaa = a, abba = bb \rangle$ from Section 4.

The first two, $\langle a, b \,|\, bab = aaa, bbbb = 1 \rangle$ and $\langle a, b \,|\, aaaa = 1, abbba = b \rangle$, differ from these finite monoids only by a single letter:

100 elements:	$\langle a, b \,\|\, bab = aa, bbbb = 1 \rangle$	$\langle a, b \,\|\, aaaa = 1, abba = b \rangle$
147 elements:	$\langle a, b \,\|\, bab = aaa, bbb = 1 \rangle$	$\langle a, b \,\|\, aaa = 1, abbba = b \rangle$
1083 elements:	$\langle a, b \,\|\, aaa = 1, bbbb = aba \rangle$	$\langle a, b \,\|\, aaa = 1, abbbba = b \rangle$

Perhaps they are large finite monoids too, and we just need to run more iterations of Knuth-Bendix completion before giving up? We can rule out this possibility.

First, for each monoid, we consider the *group* with the same presentation. This is called the *universal enveloping group* of the monoid. The universal enveloping group of $\langle a, b \,|\, bab = aaa, bbbb = 1 \rangle$ can be presented as follows, by adding a letter c to act as the inverse of a:

$$\langle a, b, c \,|\, bab = aaa, bbbb = 1, ca = 1, ac = 1 \rangle$$

This group has a finite complete presentation, with the recursive path order $a \wr c \wr b$:

$$\begin{array}{llll} ca & \Rightarrow 1 & bab & \Rightarrow aaa \\ ac & \Rightarrow 1 & abaab & \Rightarrow bbaaaaaaa \\ aaaab & \Rightarrow baaaa & aaabb & \Rightarrow baabccc \\ cb & \Rightarrow aaabcccc & bbb & \Rightarrow abccc \end{array}$$

The presented group is infinite. For example, a^n is irreducible for all $n \geq 0$.

We can repeat this with $\langle a, b \,|\, aaaa = 1, abbba = b \rangle$. We add a letter c to act as the inverse of b, which yields a group with same finite complete presentation as above, except with the renaming $a \mapsto c$, $b \mapsto a$, $c \mapsto b$.

Now, we apply the next lemma, due to Mastodon user `typeswitch` [26]:

Lemma 5.1. Let $\langle A \,|\, R \rangle$ be a monoid presentation, let M be the presented monoid, and let G be the universal enveloping group of M, so the group with the same presentation. If M is a finite monoid, then G is a finite group.

Proof. Suppose that $x \in A$. Since M is finite, there exist positive integers $m < n$ such that $x^m \Leftrightarrow x^n$. It follows that $x^m \Leftrightarrow x^n$ as elements of G, but furthermore, $x^{n-m} \Leftrightarrow 1$ in G. Multiplying both sides by x^{-1} then gives us that $x^{n-m-1} \Leftrightarrow x^{-1}$ in G, for all $x \in A$.

Now, consider the canonical mapping that sends each monoid generator of M to the corresponding group generator of G, and extend this mapping to a monoid homomorphism from M to G. We just showed that each element of G can be written with positive powers of generators only, so this homomorphism is surjective, and in particular, G is finite. □

Thus, we conclude that our unsolved monoids cannot be finite, so we ask:

Question 2. Do $\langle a, b \,|\, bab = aaa, bbbb = 1 \rangle$ and $\langle a, b \,|\, aaaa = 1, abbba = b \rangle$ admit finite complete presentations? Are they isomorphic?

Three Generators

Finally, with a small change to the program, we can enumerate presentations with *three* generators and two relations. Unlike the two generator case, this enumeration does not contain any finite monoids—it takes at least three relations to present a finite monoid on three generators. The first hard instance has length 6:

$$\langle a, b, c \,|\, ba = ac, cb = 1 \rangle$$

This remarkable monoid can be viewed as a generalization of the *bicyclic monoid*, $\langle b, c \,|\, cb = 1 \rangle$. Let us rename the generators as follows: $a \mapsto \texttt{\$}$, $b \mapsto \texttt{)}$, $c \mapsto \texttt{(}$. The defining relations become $\texttt{)\$} \Leftrightarrow \texttt{\$(}$ and $\texttt{()} \Leftrightarrow 1$. It is also easy to see that:

$$\texttt{(\$} \Leftrightarrow \texttt{(\$()} \Leftrightarrow \texttt{()\$)} \Leftrightarrow \texttt{\$)}$$

This leads to an infinite complete presentation—here the dollars collect on the left:

$$\begin{array}{ll} \texttt{()} & \Rightarrow 1 \\ \texttt{)\$} & \Rightarrow \texttt{\$(} \\ \texttt{(\$} & \Rightarrow \texttt{\$)} \\ \texttt{\$)}^n\texttt{)(} & \Rightarrow \texttt{\$)}^n \quad \text{for all } n \geq 0 \end{array}$$

The irreducible words are precisely $\texttt{)}^m\texttt{(}^n$, $\texttt{\$}^m\texttt{(}^n$, and $\texttt{\$}^m\texttt{)}^n$, where m, $n \geq 0$.

Conjecture 1. The monoid presented by $\langle a, b, c \,|\, ba = ac, cb = 1 \rangle$ does not admit a finite complete presentation.

References

[1] E. L. Post, "Recursive unsolvability of a problem of Thue," *The Journal of Symbolic Logic*, vol. 12, no. 1, p. 1–11, 1947. https://www.wolframscience.com/prizes/tm23/images/Post2.pdf

[2] F. Otto and Y. Kobayashi, "Properties of monoids that are presented by finite convergent string-rewriting systems — a survey," in *Advances in Algorithms, Languages, and Complexity*. Springer, 1997, pp. 225–266. https://static.aminer.org/pdf/PDF/000/066/293/properties_of_monoids_that_are_presented_by_finite_convergent_string.pdf

[3] D. E. Knuth and P. B. Bendix, "Simple word problems in universal algebras," in *Automation of Reasoning: 2: Classical Papers on Computational Logic 1967–1970*. Springer, 1983, pp. 342–376. https://www.semanticscholar.org/paper/Simple-Word-Problems-in-Universal-Algebras-Knuth-Bendix/94877bdf8313565b90758a5e664764139857b358

[4] G. Huet, "A complete proof of correctness of the Knuth-Bendix completion algorithm," *Journal of Computer and System Sciences*, vol. 23, no. 1, pp. 11–21, 1981. https://www.sciencedirect.com/science/article/pii/0022000081900027

[5] R. V. Book, "Thue systems as rewriting systems," *Journal of Symbolic Computation*, vol. 3, no. 1, pp. 39–68, 1987. https://www.sciencedirect.com/science/article/pii/S0747717187800214

[6] D. Kapur and P. Narendran, "A finite Thue system with decidable word problem and without equivalent finite canonical system," *Theoretical Computer Science*, vol. 35, pp. 337–344, 1985. https://www.sciencedirect.com/science/article/pii/0304397585900234

[7] F. Otto, "Finite complete rewriting systems for the Jantzen monoid and the Greendlinger group," *Theoretical Computer Science*, vol. 32, no. 3, pp. 249–260, 1984. https://www.sciencedirect.com/science/article/pii/0304397584900446

[8] G. S. Tseitin, "An associative calculus with an insoluble problem of equivalence," in *Problems of the constructive direction in mathematics. Part 1*. Moscow–Leningrad: Acad. Sci. USSR, 1958, vol. 52, pp. 172–189. https://www.mathnet.ru/php/archive.phtml?wshow=paper&jrnid=tm&paperid=1317&option_lang=eng

[9] C.-F. Nyberg-Brodda, "G. S. Tseytin's seven-relation semigroup with undecidable word problem," 2024. https://arxiv.org/abs/2401.11757

[10] C. C. Squier, "Word problems and a homological finiteness condition for monoids," *Journal of Pure and Applied Algebra*, vol. 49, no. 1, pp. 201–217, 1987. https://www.sciencedirect.com/science/article/pii/0022404987901290

[11] C. C. Squier, F. Otto, and Y. Kobayashi, "A finiteness condition for rewriting systems," *Theoretical Computer Science*, vol. 131, no. 2, pp. 271–294, 1994. https://www.sciencedirect.com/science/article/pii/0304397594901759

[12] R. Cremanns and F. Otto, "Finite derivation type implies the homological finiteness condition FP3," *Journal of Symbolic Computation*, vol. 18, no. 2, pp. 91–112, 1994. https://www.sciencedirect.com/science/article/pii/S074771718471039X

[13] M. Katsura and Y. Kobayashi, "Constructing finitely presented monoids which have no finite complete presentation," *Semigroup Forum*, vol. 54, pp. 292–302, 1997. https://link.springer.com/article/10.1007/BF02676612

[14] Y. Lafont and A. Prouté, "Church-Rosser property and homology of monoids," *Mathematical Structures in Computer Science*, vol. 1, pp. 297 – 326, 1991. https://www.irif.fr/~mellies/mpri/mpri-ens/articles/lafont-proute-church-rosser-property-and-homology-of-monoids.pdf

[15] A. J. Cain, R. D. Gray, and A. Malheiro, "On finite complete rewriting systems, finite derivation type, and automaticity for homogeneous monoids," *Information and Computation*, vol. 255, pp. 68–93, 2017. https://www.sciencedirect.com/science/article/pii/S0890540117300937

[16] Y. Kobayashi, "Finite homotopy bases of one-relator monoids," *Journal of Algebra*, vol. 229, no. 2, pp. 547–569, 2000. https://www.sciencedirect.com/science/article/pii/S0021869399982510

[17] C.-F. Nyberg-Brodda, "The word problem for one-relation monoids: a survey," *Semigroup Forum*, vol. 103, no. 2, pp. 297–355, 2021. https://link.springer.com/article/10.1007/s00233-021-10216-8

[18] R. Book and F. Otto, *String-Rewriting Systems*, ser. Monographs in Computer Science. Springer, 2012. https://link.springer.com/book/10.1007/978-1-4613-9771-7

[19] M. H. A. Newman, "On theories with a combinatorial definition of equivalence," *Annals of Mathematics*, vol. 43, no. 2, pp. 223–243, 1942. http://www.ens-lyon.fr/LIP/REWRITING/TERMINATION/NEWMAN/Newman.pdf

[20] Y. Kobayashi, "Homotopy Reduction Systems for Monoid Presentations II: The Guba—Sapir Reduction and Homotopy Modules," in *Algorithmic Problems in Groups and Semigroups*. Boston, MA: Birkhäuser Boston, 2000, pp. 143–159.

[21] B. Grechuk, *Polynomial Diophantine Equations: A Systematic Approach*. Springer International Publishing, 2024. https://link.springer.com/book/10.1007/978-3-031-62949-5

[22] R. McNaughton, "The finiteness of finitely presented monoids," *Theoretical Computer Science*, vol. 204, no. 1, pp. 169–182, 1998. https://www.sciencedirect.com/science/article/pii/S0304397598000383

[23] J. Berstel, *Transductions and Context-Free Languages*. Teubner Verlag, 1979. https://www-igm.univ-mlv.fr/~berstel/LivreTransductions/LivreTransductions.html

[24] C. C. Sims, *Computation with Finitely Presented Groups*, ser. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 1994. https://www.cambridge.org/us/universitypress/subjects/mathematics/algebra/computation-finitely-presented-groups

[25] J. Pedersen, "Morphocompletion for one-relation monoids," in *Rewriting Techniques and Applications*, N. Dershowitz, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 1989, pp. 574–578.

[26] "Mastodon post," 2025. https://gamedev.lgbt/@typeswitch/114915626980225127