Rational subsets of Baumslag-Solitar groups

Michaël Cadilhac

DePaul University, Chicago, IL, USA

michael@cadilhac.name

Dmitry Chistikov

Centre for Discrete Mathematics and its Applications (DIMAP) & Department of Computer Science, University of Warwick, United Kingdom d.chistikov@warwick.ac.uk

Georg Zetzsche

Max Planck Institute for Software Systems (MPI-SWS), Germany georg@mpi-sws.org

- Abstract

We consider the rational subset membership problem for Baumslag-Solitar groups. These groups form a prominent class in the area of algorithmic group theory, and they were recently identified as an obstacle for understanding the rational subsets of $GL(2, \mathbb{Q})$.

We show that rational subset membership for Baumslag-Solitar groups $\mathrm{BS}(1,q)$ with $q\geq 2$ is decidable and PSPACE-complete. To this end, we introduce a word representation of the elements of $\mathrm{BS}(1,q)$: their pointed expansion (PE), an annotated q-ary expansion. Seeing subsets of $\mathrm{BS}(1,q)$ as word languages, this leads to a natural notion of PE-regular subsets of $\mathrm{BS}(1,q)$: these are the subsets of $\mathrm{BS}(1,q)$ whose sets of PE are regular languages. Our proof shows that every rational subset of $\mathrm{BS}(1,q)$ is PE-regular.

Since the class of PE-regular subsets of $\mathrm{BS}(1,q)$ is well-equipped with closure properties, we obtain further applications of these results. Our results imply that (i) emptiness of Boolean combinations of rational subsets is decidable, (ii) membership to each fixed rational subset of $\mathrm{BS}(1,q)$ is decidable in logarithmic space, and (iii) it is decidable whether a given rational subset is recognizable. In particular, it is decidable whether a given finitely generated subgroup of $\mathrm{BS}(1,q)$ has finite index.

2012 ACM Subject Classification Theory of computation \rightarrow Problems, reductions and completeness; Theory of computation \rightarrow Formal languages and automata theory; Theory of computation \rightarrow Models of computation

Keywords and phrases Rational subsets, Baumslag-Solitar groups, decidability, regular languages, pointed expansion

Acknowledgements The research for this work was carried out in part at the Autobóz Research Camp in 2019 in Firbush, Scotland. The authors are very grateful to Piotr Hofman for comments that led to important insights and to Michael Blondin as well as to all other participants for discussions.

M. Cadilhac, D. Chistikov, and G. Zetzsche

_	— C	ontents — — — — — — — — — — — — — — — — — — —	
1	Introduction		1
2			3 5
3			
4	Closure properties		6
	4.1	The PE-regular subsets of $BS(1,q)$ form a Boolean algebra	6
	4.2	The rational subsets of BS $(1, q)$ are not closed under intersection	6
	4.3	The PE-regular subsets of $BS(1,q)$ are not closed under iteration	7
5	Every rational subset of $BS(1, q)$ is effectively PE-regular		8
	5.1	Any run is equivalent to a thin run augmented with thin returning-left cycles	10
	5.2	Intermezzo: reflecting on Corollary 5.3	11
	5.3	Sets of thin runs are PE-regular	12
	5.4	Iterations of returning-left thin cycles are PE-regular	14
	5.5	Wrapping up: Proof of Theorem 3.1	14
6	Complexity		15
	6.1	PSPACE-hardness	15
	6.2	PSPACE membership	17
7	Recognizability 2		21

1 Introduction

Subsets of groups Regular languages are an extremely versatile tool in algorithmics on sets of finite words. This is mainly due to two reasons. First, they are robust in terms of representations and closure properties: They can be described by finite automata, by recognizing morphisms, and by monadic second-order logic and they are closed under Boolean and an abundance of other operations. Second, many properties (such as emptiness) are easily decidable using finite automata.

Given this success, there have been several attempts to develop an analogous notion for subsets of (infinite, finitely generated) groups. Adapting the notion of recognizing morphism yields $recognizable\ subsets$ of a group G. They are closed under Boolean operations, and problems such as membership or emptiness are decidable. However, since they are merely unions of cosets of finite-index normal subgroups, their expressiveness is severely limited.

Another notion is that of *rational subsets*, which transfer (non-deterministic) finite automata to groups. Starting with pioneering work by Benois [6] in 1969, they have matured into an important tool in group theory. Rational subsets are quite expressive: They include finitely generated submonoids and are closed under (finite) union, pointwise product, and Kleene star. Moreover, they have been applied successfully to solving equations in groups [12, 10], as well as in other settings [2, 37].

The high expressiveness of rational subsets comes at the cost of undecidability of decision problems for many groups. The most fundamental one is the membership problem for rational subsets: Given a rational subset R of a group G and an element $g \in G$, does g belong to R? Understanding for which groups this problem is decidable received significant attention over the last two decades, see [27] for a survey. Unfortunately, the rational subsets do not quite reach the level of robustness of regular languages. In general, the class of rational subsets of a group is not closed under Boolean operations, and the papers [28, 4] study for which groups the rational subsets form a Boolean algebra.

Baumslag-Solitar groups A prominent class of groups is that of Baumslag-Solitar groups BS(p,q). For each $p,q \in \mathbb{N}$, the group is defined as $BS(p,q) = \langle a,t \mid ta^pt^{-1} = a^q \rangle$. They were introduced in 1962 by Baumslag and Solitar to provide an example of a two-generator one-relator group that is non-Hopfian. They recently came into focus from the algorithmic perspective in a paper by Kharlampovich, López, and Miasnikov [24], which shows that solvability of equations is decidable in BS(1,q). They have also been studied from several other perspectives, such as the decidability and complexity of the word problem [31, 15, 38], the conjugacy problem [15, 38], tiling problems [1], and computing normal forms [14, 19, 18].

More specifically to our setting, the Baumslag-Solitar groups have recently been identified by Diekert, Potapov, and Semukhin [16] as a stumbling block in solving rational subset membership in the group $GL(2,\mathbb{Q})$, that is, the group of invertible 2×2 matrices over \mathbb{Q} . They show that any subgroup of $GL(2,\mathbb{Q})$ containing $GL(2,\mathbb{Z})$ is either of the form $GL(2,\mathbb{Z})\times\mathbb{Z}^k$ for $k\geq 1$ or contains BS(1,q) as a subgroup for some $q\geq 2$. Rational subset membership for $GL(2,\mathbb{Z})\times\mathbb{Z}^k$ is today a matter of standard arguments [27], because $GL(2,\mathbb{Z})$ is virtually free. Therefore, making significant progress towards decidability in larger subgroups requires understanding rational subsets of BS(1,q).

One can represent the elements of BS(1, q) as pairs (r, m), where r is a number in $\mathbb{Z}\left[\frac{1}{q}\right]$, say $r = \pm \sum_{i=-n}^{n} a_i q^i$ for $a_{-n}, a_{-n+1}, \ldots, a_n \in \{0, \ldots, q-1\}$, and $m \in \mathbb{Z}$. Here, one can

¹ $\mathbb{Z}[\frac{1}{q}]$ denotes (the additive group of) the smallest subring of $(\mathbb{Q}, +, \cdot)$ containing \mathbb{Z} and 1/q; as a set, it

2 Rational subsets of Baumslag-Solitar groups

think of m as a cursor pointing to a position in the q-ary expansion $a_nq^n+\cdots+a_{-n}q^{-n}$. Then the action of the generators of BS(1, q) is as follows. Multiplication by t or t^{-1} moves the cursor to the left or the right, respectively. Multiplication by a adds q^m ; likewise, multiplication by a^{-1} subtracts q^m . Thus, from an automata-theoretic perspective, one can view the rational subset membership problem as the reachability problem for an extended version of one-counter automata. Instead of storing a natural number, such an automaton stores a number $r \in \mathbb{Z}[\frac{1}{q}]$. Moreover, instead of instructions "increment by 1" and "decrement by 1", it has an additional \mathbb{Z} -counter m that determines the value to be added in the next update. Then, performing "increment" on r will add q^m and "decrement" on r will subtract q^m . The \mathbb{Z} -counter m supports the classical "increment" and "decrement" instructions.

Contribution Our first main contribution is to show is that for each group BS(1,q), the rational subset membership problem is decidable and PSPACE-complete. To this end, we show that each rational subset can be represented by a regular language of finite words that encode elements of BS(1,q) in the natural way: For (r,m) as above, we encode each digit a_i by a letter; and we decorate the digits at position 0 and at position m. We call this encoding the pointed expansion (PE) of (r,m). This leads to a natural notion of subsets of BS(1,q), which we call PE-regular. We regard the introduction of this notion as the second main contribution of this work.

The class of PE-regular subsets of BS(1,q) has several properties that make them a promising tool for decision procedures for BS(1,q): First, our proof shows that it effectively includes the large class of rational subsets, in particular any finitely generated submonoid. Second, they form an effective Boolean algebra. Third, due to them being regular languages of words, they inherit many algorithmic tools from the setting of free monoids. We apply these properties to obtain three applications of our main results.

- 1. Membership in each fixed rational subset can be decided in logarithmic space.
- 2. We show that it is decidable whether a given PE-regular subset (and thus a given rational subset) is recognizable. Recognizability of rational subsets is rarely known to be decidable for groups: The only examples known to the authors are free groups, for which decidability was shown by Sénizergues [34] (and simplified by Silva [36]) and free abelian groups (this follows from [21, Theorem 3.1]). Since (i) finitely generated subgroups are rational subsets and (ii) a subgroup of any group G is recognizable if and only if it has finite index in G, our result implies that it is decidable whether a given finitely generated subgroup of BS(1,q) has finite index. Studying decidability of this finite index problem in groups was recently proposed by Kapovich [13, Section 4.3].
- 3. Our results imply that emptiness of Boolean combinations (hence inclusion, equality, etc.) of rational subsets is decidable. (We also show that the rational subsets of BS(1, q) are not closed under intersection.) This is a strong decidability property that already fails for groups as simple as $F_2 \times \mathbb{Z}$ (this follows from [22, Theorem 6.3]), where F_2 is the free group over two generators, and hence for $GL(2,\mathbb{Z}) \times \mathbb{Z}^k$, $k \geq 1$.

Finally, we remark that since $\mathrm{BS}(1,q)$ is isomorphic to the group of all matrices $\binom{q^m}{0} \binom{r}{1}$ for $m \in \mathbb{Z}$ and $r \in \mathbb{Z}[\frac{1}{q}]$, our results can be interpreted as solving the rational subset membership problem for this subgroup of $\mathrm{GL}(2,\mathbb{Q})$.

Related work It is well-known that membership in a given finitely generated subgroup, called the *generalized word problem* of BS(1,q), is decidable. This is due to a general result

consists of all rational numbers of the form $n \cdot q^j$, $n, j \in \mathbb{Z}$.

of Romanovskii, who showed in [32] and [33] that solvable groups of derived length two have a decidable generalized word problem (it is an easy exercise to show that BS(1,q) is solvable of derived length two for each $q \in \mathbb{N}$).

Another restricted version of rational subset membership is the knapsack problem, which was introduced by Myasnikov, Nikolaev, and Ushakov [30]. Here, one is given group elements g_1, \ldots, g_k, g and is asked whether there exist $x_1, \ldots, x_k \in \mathbb{N}$ with $g_1^{x_1} \cdots g_k^{x_k} = g$. A recent paper on the knapsack problem in Baumslag-Solitar groups by Dudkin and Treyer [17] left open whether the knapsack problem is decidable in BS(1,q) for $q \geq 2$. This was settled very recently in [29], where one expresses solvability of $g_1^{x_1} \cdots g_k^{x_k} = g$ in a variant of Büchi arithmetic. A slight extension of that proof yields a regular language as above for the set $S = \{g_1^{x_1} \cdots g_k^{x_k} \mid x_1, \ldots, x_k \in \mathbb{N}\}$. Note that each element g_i moves the cursor either to the left (i.e. increases m), to the right (i.e. decreases m), or not at all. Thus, in a product $g_1^{x_1} \cdots g_k^{x_k}$, the cursor direction is reversed at most k-1 times. The challenge of our translation from rational subsets to PE-regular subsets is to capture products where the cursor changes direction an unbounded number of times.

Finally, closely related to rational subsets, there is another approach to group-theoretic problems via automata: One can represent finitely generated subgroups of free groups using *Stallings graphs*. Due to the special setting of free groups, they behave in many ways similar to automata over words and are thus useful for decision procedures [23]. Stallings graphs have recently been extended to semidirect products of free groups and free abelian groups by Delgado [11]. However, this does not include products $\mathbb{Z}[\frac{1}{g}] \times \mathbb{Z}$ and is restricted to subgroups.

2 Basic notions

Automata, rational subsets, and regular languages Since we work with automata over finite words and over groups, we define automata over a general monoid M. A subset $S \subseteq M$ is recognizable if there is a finite monoid F and a morphism $\varphi \colon M \to F$ such that $S = \varphi^{-1}(\varphi(S))$. If M is a group, one can equivalently require F to be a finite group.

For a subset $S \subseteq M$, we write $\langle S \rangle$ or S^* for the submonoid generated by S, i.e. the set of elements that can be written as a (possibly empty) product of elements of S. In particular, the neutral element $1 \in M$ always belongs to $\langle S \rangle = S^*$. A generating set is a subset $\Sigma \subseteq M$ such that $M = \langle \Sigma \rangle$. We say that M is finitely generated (f.g.) if it has a finite generating set. Suppose M is finitely generated and fix a finite generating set Σ . An automaton over M is a tuple $A = (Q, \Sigma, E, q_0, q_f)$, where Q is a finite set of states, $E \subseteq Q \times \Sigma \times Q$ is a finite set of edges, $q_0 \in Q$ is its initial state, and $q_f \in Q$ is its final state. A run (in A) is a sequence $\rho = (p_0, a_1, p_1) \cdots (p_{m-1}, a_m, p_m)$, where $(p_{i-1}, a_i, p_i) \in E$ for $i \in [1, m]$. It is accepting if $p_0 = q_0$ and $p_m = q_f$. By $[\rho]$, we denote the production of ρ , that is, the element $a_1 \cdots a_m \in M$. Two runs are equivalent if they start in the same state, end in the same state, and have the same production. For a set of runs P, we denote $[P] = \{[\rho] \mid \rho \in P\}$.

The subset accepted by \mathcal{A} is $\mathsf{L}(\mathcal{A}) = \{[\rho] \mid \rho \text{ is an accepting run in } \mathcal{A}\}$. A subset $R \subseteq M$ is called rational if it is accepted by some automaton over M. It is a standard fact that the family of rational subsets of M does not depend on the chosen generating set Σ . Rational subsets of a free monoid Γ^* for some alphabet Γ are also called $regular\ languages$. If $M = \Gamma^* \times \Delta^*$ for alphabets Γ, Δ , then rational subsets of M are also called $rational\ transductions$. If $T \subseteq \Gamma^* \times \Delta^*$ and $L \subseteq \Gamma^*$, then we set $TL = \{v \in \Delta^* \mid \exists u \in L : (u,v) \in T\}$. It is well-known that if $L \subseteq \Gamma^*$ is regular and $T \subseteq \Gamma^* \times \Delta^*$ is rational, then TL is regular as well [7].

Baumslag-Solitar groups The Baumslag-Solitar groups are the groups $\mathrm{BS}(p,q)$ for $p,q\in\mathbb{N}$, where $\mathrm{BS}(p,q)=\langle a,t\mid ta^pt^{-1}=a^q\rangle$. They were introduced in 1962 by Baumslag and Solitar [3] to provide an example of a non-Hopfian group with two generators and one defining relation. In this paper, we focus on the case p=1. In this case, there is a well-known isomorphism $\mathrm{BS}(1,q)\cong\mathbb{Z}[\frac{1}{q}]\rtimes\mathbb{Z}$ and we will identify the two groups. Here, $\mathbb{Z}[\frac{1}{q}]$ is the additive group of number nq^i with $n,i\in\mathbb{Z}$, and \rtimes denotes semidirect product. Building this semidirect product requires us to specify an automorphism φ_m of $\mathbb{Z}[\frac{1}{q}]$ for each $m\in\mathbb{Z}$, which is given by $\varphi_m(nq^i)=q^m\cdot nq^i$.

For readers not familiar with semidirect products, we give an alternative self-contained definition of $\mathbb{Z}[\frac{1}{q}] \rtimes \mathbb{Z}$. The elements of this group are pairs (r, m), where $r \in \mathbb{Z}[\frac{1}{q}]$ and $m \in \mathbb{Z}$. The multiplication is defined as

$$(r,m)(r',m') = (r+q^m \cdot r', m+m').$$

We think of an element (r,m) as representing a number r in $\mathbb{Z}[\frac{1}{q}]$ together with a cursor m to a position in the q-ary expansion of r. Multiplying an element (r,m) by the pair (1,0) from the right means adding 1 at the position in r given by m, hence adding q^m to r and leaving the cursor unchanged: we have $(r,m)(1,0)=(r+q^m,m)$. Multiplying by (0,1) moves the cursor one position to the left: (r,m)(0,1)=(r,m+1). It is easy to see that $\mathbb{Z}[\frac{1}{q}] \times \mathbb{Z}$ is generated by the set $\{(1,0),(-1,0),(0,1),(0,-1)\}$. The isomorphism $\mathrm{BS}(1,q) \xrightarrow{\sim} \mathbb{Z}[\frac{1}{q}] \times \mathbb{Z}$ mentioned above maps a to (1,0) and t to (0,1). Since we identify $\mathrm{BS}(1,q)$ and $\mathbb{Z}[\frac{1}{q}] \times \mathbb{Z}$, we will have a=(1,0) and t=(0,1). In particular, a can be thought of as "add"/"increment", and t as "move". We regard elements of the subgroup $\mathbb{Z}[\frac{1}{q}] \times \{0\}$ of $\mathrm{BS}(1,q)$ as elements of $\mathbb{Z}[\frac{1}{q}]$, i.e., integers or rational fractions with denominator q^i , $i \geq 1$.

Rational subset membership Unless specified otherwise, automata over BS(1,q) will use the generating set $\Sigma = \{a, a^{-1}, t, t^{-1}\} = \{(1,0), (-1,0), (0,1), (0,-1)\}$. The central decision problem of this work is the *rational subset membership problem for* BS(1,q):

Given An automaton $\mathcal A$ over BS(1,q) and an element $g \in BS(1,q)$ as a word over Σ .

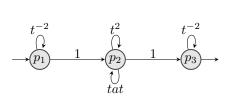
Question Does g belong to $L(\mathcal A)$?

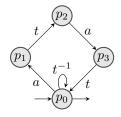
Automata over BS(1, q) In the following definitions, let $\mathcal{A} = (Q, \Sigma, E, q_0, q_f)$ be an automaton over BS(1, q). For a run ρ of \mathcal{A} , recall that $[\rho] \in \mathbb{Z}[\frac{1}{q}] \rtimes \mathbb{Z}$ is the production of ρ . Moreover, if $[\rho] = (r, m)$ with $r \in \mathbb{Z}[\frac{1}{q}]$ and $m \in \mathbb{Z}$, then we define $\mathsf{pos}(\rho) = m$, and call this the final position of ρ . More generally, the position at a particular point in ρ is the final position of the corresponding prefix of ρ . By $\mathsf{pmax}(\rho)$, we denote the maximal value of $\mathsf{pos}(\pi)$ where π is a prefix of ρ . Analogously, $\mathsf{pmin}(\rho)$ is the minimal value of $\mathsf{pos}(\pi)$ where π is a prefix of ρ . A run ρ is returning if $\mathsf{pos}(\rho) = 0$. It is returning-left if in addition $\mathsf{pmin}(\rho) = 0$. Note that for a returning run ρ , we have $[\rho] \in \mathbb{Z}[\frac{1}{q}]$ and if ρ is returning-left, we have $[\rho] \in \mathbb{Z}$. Let $|\rho|$ be the length of the run ρ as a word over E. We will often write ρ_i assuming $\rho = \rho_1 \rho_2 \dots \rho_\ell$ where each $\rho_i \in E$ and $\ell = |\rho|$. A run is a cycle if it is returning and starts and ends in the same state. The thickness of a run ρ is defined as the greatest number of times a position is seen:

$$\mathsf{thickness}(\rho) = \max_{n \in \mathbb{Z}} |\{i \mid \mathsf{pos}(\rho_1 \cdots \rho_i) = n\}| \enspace .$$

We call a run k-thin if its thickness is at most k.

We let $\mathsf{Runs}(\mathcal{A})$ (resp. $\mathsf{Ret}(\mathcal{A})$, $\mathsf{RetL}(\mathcal{A})$) be the set of all accepting runs (resp. accepting returning runs, accepting returning-left runs) of \mathcal{A} . We add k in subscript to restrict the set





- (a) Automaton over BS(1,q) from Section 4.2.
- **(b)** Automaton over BS(1,2) from Example 5.1.

Figure 1 Example automata over BS(1, q).

to k-thin runs; for instance, $\operatorname{Ret}_k(\mathcal{A})$ is the set of k-thin returning runs. Further, we write $\operatorname{Runs}_k^{p\to p'}(\mathcal{A})$ for k-thin runs that start in p and end in p', and use the similar notations $\operatorname{Ret}_k^{p\to p'}(\mathcal{A})$ and $\operatorname{RetL}_k^{p\to p'}(\mathcal{A})$.

Seeing $\{0,\ldots,q-1\}$ as an alphabet, write Φ_q for letters from this alphabet with possibly a \bullet subscript (e.g., 0_{\bullet}), a \triangleleft superscript (e.g., $0_{\bullet}^{\triangleleft}$), or both (e.g., $0_{\bullet}^{\triangleleft}$). For $v=(r,n)\in \mathrm{BS}(1,q)$, we write $\mathrm{pe}(v)$ for its base-q pointed expansion (or just expansion) as a word in $\pm \Phi_q^*$, where the subscript \bullet and the superscript \triangleleft appear only once, the former representing the radix point, the latter indicating the value of n. That is, if $r=\sum_{i=-k_2}^{k_1}a_iq^i$, with $k_1,k_2\geq 0$, $\mathrm{pe}(v)$ is the following word:

$$\pm a_{k_1} \cdots a_1(a_0) \bullet a_{-1} \cdots a_{-k_2}$$

where \triangleleft is added to a_n . We tacitly assume a uniqueness condition: the expansion pe(v) of an element $v \in BS(1,q)$ is the shortest that abides by the definition. Expansions are read by automata in the left to right direction, i.e., from most to least significant digit.

▶ **Definition 2.1.** We say that a subset of $R \subseteq BS(1,q)$ is PE-regular, where PE stands for pointed expansion, if the word language $\{pe(v) \mid v \in R\}$ is regular.

We remark that basic properties of regular languages support the transformation of noncanonical expansions of elements BS(1, q), i.e., those with zeros on the left or right, into canonical ones, $\mathsf{pe}(v)$. Finally, recall that we identify each $r \in \mathbb{Z}[\frac{1}{q}]$ with $(r,0) \in \mathbb{Z}[\frac{1}{q}] \rtimes \mathbb{Z}$. Hence, for $r \in \mathbb{Z}[\frac{1}{q}]$, $\mathsf{pe}(r)$ is the q-ary expansion of r (with \triangleleft as an additional decoration at the radix point).

3 Main results

In this section, we list our main contributions, their proofs being deferred to later sections. Our first main result is that one can translate rational subsets into PE-regular subsets.

▶ **Theorem 3.1.** Every rational subset of BS(1,q) is effectively PE-regular.

This will be shown in Section 5. Since membership is decidable for regular languages and given $g \in BS(1,q)$ as a word over $\{a,a^{-1},t,t^{-1}\}$, one can compute pe(g), Theorem 3.1 implies that rational subset membership is decidable. Our next main result is that the problem is PSPACE-complete.

▶ **Theorem 3.2.** The rational subset membership problem for BS(1,q) is PSPACE-complete.

This is shown in Section 6. We shall also conclude that membership to each fixed rational subset is decidable in logspace.

▶ **Theorem 3.3.** For each fixed rational subset of BS(1,q), membership is decidable in logarithmic space.

The proof can also be found in Section 6. Note that, in particular, membership to each fixed subgroup of BS(1,q) is decidable in logarithmic space. Another application of Theorem 3.1 is that one can decide whether a given rational subset of BS(1,q) is recognizable.

▶ **Theorem 3.4.** Given a PE-regular subset R of BS(1,q), it is decidable whether R is recognizable.

This is shown in Section 7. Since a subgroup of any group H is recognizable if and only if it has finite index in H (see, e.g. [2, Prop. 3.2]), we obtain:

▶ Corollary 3.5. Given a f.g. subgroup of BS(1,q), it is decidable whether it has finite index.

4 Closure properties

In this section, we show some closure properties of rational and PE-regular subsets of BS(1,q). Our goal is twofold: First, give a hands-on introduction to these concepts, and second, contrast them by exhibiting structural differences between these sets.

4.1 The PE-regular subsets of BS(1, q) form a Boolean algebra

▶ Proposition 4.1. The PE-regular subsets of BS(1,q) form an effective Boolean algebra. Moreover, for PE-regular subsets $R, S \subseteq BS(1,q)$, the sets $RS = \{rs \mid r \in R, s \in S\}$ and $R^{-1} = \{r^{-1} \mid r \in R\}$ are PE-regular as well.

Proof. The first statement is due to the fact that the regular languages form an effective Boolean algebra and that the set of all pe(g) for $g \in BS(1, q)$ is regular.

It is easy to construct an automaton \mathcal{M} over $\Gamma^* \times \Gamma^* \times \Gamma^*$, for suitable Γ , that accepts the relation $T = \{(\mathsf{pe}(g), \mathsf{pe}(h), \mathsf{pe}(gh)) \mid g, h \in \mathrm{BS}(1,q)\}$: It makes sure that the radix point of the word in the second component is aligned with the cursor position of the word in the first component. Then, multiplying the two elements amounts to adding up the q-ary expansions (see also Lemma 5.5 for a more general statement). Given automata for $\mathsf{pe}(R)$ and $\mathsf{pe}(S)$, we can easily modify \mathcal{M} so as to accept $\{(\mathsf{pe}(g), \mathsf{pe}(h), \mathsf{pe}(gh)) \mid g \in R, h \in S\}$. Projecting to the third component then yields an automaton for the language $\mathsf{pe}(RS)$. A similar modification of \mathcal{M} leads to $\{(\mathsf{pe}(g), \mathsf{pe}(h), \mathsf{pe}(gh)) \mid g \in R, h \in \mathrm{BS}(1,q), \mathsf{pe}(gh) = \mathsf{pe}(1)\}$.

Together with Theorem 3.1, this implies that emptiness of Boolean combinations (hence inclusion, equality) is decidable for rational subsets. To further highlight the advantages of PE-regular subsets, we also show that the rational subsets of BS(1,q) are not closed under intersection.

4.2 The rational subsets of BS(1, q) are not closed under intersection

We present an example of rational subsets $R_1, R_2 \subseteq \mathrm{BS}(1,q)$ such that $R_1 \cap R_2$ is not rational. Let R be the rational subset accepted by the automaton in Figure 1a. In p_1 , it moves the cursor an even number of positions to the right. In p_2 , it moves an even number of positions to the left and on the way, it adds q in a subset of the even positions. In p_3 , it moves to the right again. Then R contains all elements $(r,m) \in \mathbb{Z}[\frac{1}{q}] \rtimes \mathbb{Z}$ where $r = \sum_{i \in A} q^{2i+1}$ for some finite $A \subseteq \mathbb{Z}$ and $m \in 2\mathbb{Z}$. Now consider the sets $R_1 = aR$, $R_2 = Ra$, and their intersection

 $I=R_1\cap R_2$. Then we have $(r,m)\in R_1$ if and only if $r=1+\sum_{i\in A}q^{2i+1}$ and $m\in 2\mathbb{Z}$ for some finite $A\subseteq \mathbb{Z}$. Moreover, $(r,m)\in R_2$ if and only if $r=q^m+\sum_{i\in A}q^{2i+1}$ and $m\in 2\mathbb{Z}$ for some finite $A\subseteq \mathbb{Z}$. Therefore, we have $(r,m)\in R_1\cap R_2$ if and only if $r=1+\sum_{i\in A}q^{2i+1}$ and m=0 for some finite $A\subseteq \mathbb{Z}$. Using the following lemma, we shall conclude that $I=R_1\cap R_2$ is not rational.

▶ **Lemma 4.2.** Let $R \subseteq \mathbb{Z}[\frac{1}{q}] \rtimes \mathbb{Z}$ be a rational subset. If $R \subseteq \mathbb{Z}[\frac{1}{q}] \times \{0\}$, then there is a $k \in \mathbb{N}$ with $R \subseteq \frac{1}{a^k}\mathbb{Z} \times \{0\}$.

Intuitively, this says that if all elements in a rational subset have the cursor in the origin, then its elements must have bounded precision. This can be shown using a pumping argument: If R did contain elements with high powers of q in the denominator, then the cursor must move arbitrarily far to the right, but then it can also end up to the right of the origin, which is impossible. Since $I \subseteq \mathbb{Z}[\frac{1}{q}] \times \{0\}$ contains $(1+q^{-2i+1},0)$ for any $i \in \mathbb{N}$, it cannot be rational.

For the detailed proof of Lemma 4.2, it is more convenient to argue with the well-known observation that an automaton that accepts a fixed element has to encode the element read so far in its state. Let us make this formal. If $\mathcal{A}=(Q,\Sigma,E,q_0,F)$ is an automaton over a group G, then a state evaluation is a map $\eta\colon Q\to G$ such that $\eta(q_0)=1$ and for every edge $(p,g,p')\in E$, we have $\eta(p')=\eta(p)g$. Hence, a state evaluation assigns to each state p a fixed group element $\eta(p)$ such that on any path from q_0 to p, \mathcal{A} reads $\eta(p)$. An automaton is called trim if (i) every state is reachable from an initial state and (ii) from every state, one can reach a final state.

▶ **Lemma 4.3.** Let A be a trim automaton over a group G that accepts the set $\{1\}$. Then A admits a state evaluation.

Proof. Since \mathcal{A} is trim, we can choose $\eta: Q \to G$ such that for every $p \in Q$, there is a run from q_0 to p in \mathcal{A} that reads $\eta(p)$.

The fact that \mathcal{A} accepts $\{1\}$ implies that there is only one such η : Suppose ρ_1 , ρ_2 are runs from q_0 to p and ρ is a run from p to a final state. Then since \mathcal{A} accepts $\{1\}$, we have $[\rho_1][\rho] = 1 = [\rho_2][\rho]$ and thus $[\rho_1] = [\rho_2]$. Hence, η is uniquely determined.

This implies that η is a state evaluation: We must have $\eta(q_0) = 1$, because of uniqueness of η . Moreover, if there is an edge (p, g, p'), then we can pick a run ρ from q_0 to p and by uniqueness of η , we have $\eta(p') = [\rho]g = \eta(p)g$.

Using Lemma 4.3, we are ready to prove Lemma 4.2.

Proof. Suppose \mathcal{A} is an automaton over $\mathbb{Z}[\frac{1}{q}] \times \mathbb{Z}$ that accepts a subset of $\mathbb{Z}[\frac{1}{q}] \times \{0\}$. Without loss of generality, we may assume that \mathcal{A} is trim and every edge has a label in $\{t, t^{-1}, a, a^{-1}\}$. Consider the automaton \mathcal{A}' obtained from \mathcal{A} by projecting to the right component. Then \mathcal{A}' is a trim automaton over \mathbb{Z} that accepts $\{0\}$. According to Lemma 4.3, \mathcal{A}' admits a state evaluation $\eta: Q \to \mathbb{Z}$. Since Q is finite, the image of η is included in some interval [-k, k].

This implies that for any state p of \mathcal{A} , any element (r,m) read on a path from q_0 to p satisfies $m \in [-k, k]$. Therefore, every edge labeled $a^{\pm 1}$ adds a number $s = \pm q^m$ with $m \in [-k, k]$ to the left component. Since in this case $s \in \frac{1}{q^k}\mathbb{Z}$, the lemma follows.

4.3 The PE-regular subsets of BS(1, q) are not closed under iteration

The subset $A = \{(1 + 2^{-i}, 0) \mid i \geq 1\}$ of BS(1, 2) is PE-regular, because $pe(A) = 1^{\triangleleft}_{\bullet}0^*1$ is a regular language. Let us now prove that the set A^* is indeed not PE-regular. We begin with an auxiliary lemma.

▶ Lemma 4.4. Suppose $k, m \ge 0$ and $1 \le d_1 \le \cdots \le d_k$ and $1 \le e_1 < e_2 < \cdots < e_\ell$ with

$$\sum_{i=1}^{k} (1 + 2^{-d_i}) = m + \sum_{i=1}^{\ell} 2^{-e_i}$$
 (1)

Then $m \geq \ell$.

Proof. We prove $m \ge k$ and $k \ge \ell$. We begin with $m \ge k$. Let s be the value of the two sums. Then clearly $k \le s$ and s < m+1, hence $k \le m+1$. Since both k and m are integers, it is impossible that k > m. Thus $k \le m$.

The inequality $k \ge \ell$ follows by induction on k. Suppose that Equation (1) holds and we add $1 + 2^{-d_{k+1}}$. We distinguish two cases:

- If in the binary expansion on the right, there is no digit $2^{-d_{k+1}}$, then the new binary expansion gains one 1 digit and hence ℓ increases by one.
- If there already is a digit at $2^{-d_{k+1}}$, then the new binary expansion is obtained by flipping some $r \ge 1$ digits from 1 to 0 and flipping one 0 into a 1. Hence, ℓ drops by r and rises by ≤ 1 .

In any case, the value for ℓ rises by at most one. This proves $k \geq \ell$.

We regard $\mathbb{Z}[\frac{1}{q}]$ as a subset of $\mathbb{Z}[\frac{1}{q}] \rtimes \mathbb{Z}$ by identifying $r \in \mathbb{Z}[\frac{1}{q}]$ with $(r,0) \in \mathbb{Z}[\frac{1}{q}] \rtimes \mathbb{Z}$. Then in particular for $m \in \mathbb{Z}$, $\mathsf{pe}(m) \in \pm \{0,\ldots,q-1\}^*\{0^{\triangleleft}_{\bullet},\ldots,(q-1)^{\triangleleft}_{\bullet}\}$ is the q-ary expansion of m, with the additional $^{\triangleleft}$ and $_{\bullet}$ at the right-most digit.

▶ **Lemma 4.5.** Let $n \in \mathbb{N}$. Then n is the smallest number $m \in \mathbb{N}$ with $pe(m) \cdot 1^n \in pe(A^*)$.

Proof. Since $n+2^{-1}+\cdots+2^{-n}=\sum_{i=1}^n(1+2^{-i})$ clearly belongs to A^* , we have $\operatorname{pe}(n)\cdot 1^n\in\operatorname{pe}(A^*)$. Now suppose $\operatorname{pe}(m)\cdot 1^n\in\operatorname{pe}(A^*)$. Then we have

$$m + 2^{-1} + \dots + 2^{-n} = \sum_{i=1}^{k} (1 + 2^{-d_i})$$

for some $k \geq 0$ and some $1 \leq d_1 \leq d_2 \leq \cdots \leq d_k$. By Lemma 4.4, this implies $m \geq n$.

Now Lemma 4.5 allows us to show that $\operatorname{pe}(A^*)$ is not regular. Recall that for a language $L\subseteq \Gamma^*$, a right quotient is a set of the form $Lu^{-1}:=\{v\in \Gamma^*\mid vu\in L\}$. Since a regular language has finite syntactic monoids (see, e.g. [7]), it has only finitely many right quotients. Suppose $\operatorname{pe}(A^*)$ is regular. For each $n\in\mathbb{N}$, consider the right quotient $Q_n=\operatorname{pe}(A^*)(1^n)^{-1}$. Then according to Lemma 4.5, for each $n\in\mathbb{N}$, n is the smallest number m with $\operatorname{pe}(m)\in Q_n\cap\operatorname{pe}(\mathbb{Z})$. Thus, the sets Q_0,Q_1,Q_2,\ldots are pairwise distinct, contradicting the fact that $\operatorname{pe}(A^*)$ has only finitely many right quotients.

5 Every rational subset of BS(1, q) is effectively PE-regular

In this section, we prove Theorem 3.1. We first illustrate our approach on an example.

▶ Example 5.1. Consider the automaton over BS(1,2) in Figure 1b. In its only initial and final state p_0 , it has a choice of two operations: (i) move the cursor one position to the right (i.e. multiplication by t^{-1}) or (ii) perform the increment on two neighbouring cells and stop one position left of them (i.e. multiplication by atat). The automaton can perform these operations arbitrarily many times in any order.

We shall prove that the automaton accepts

$$R = \{(3n \cdot 2^{m-2k}, m) \mid n \in \mathbb{N}, \ k \in \mathbb{N}, \ m \in \mathbb{Z}, \ 0 \ge m - 2k, \ 3n \cdot 2^{m-2k} \ge f(m, k)\} \ ,$$

where

$$f(m,k) = \sum_{i=1}^{k} 3 \cdot 2^{m-2i} = \sum_{j=m-2k}^{m-1} 2^{j} = 2^{m} - 2^{m-2k}$$
.

The language pe(R) is regular. Indeed, note that the number f(m, k) has a particularly simple binary representation. A pointed expansion of (r, m) belongs to pe(R) if there is a position $m - 2k \le 0$ such that reading the digits left of position m - 2k yields a number (namely 3n) that (a) is divisible by 3 and (b) lies above a bound with a simple binary expansion.

Let us now prove that the automaton accepts R. Let ρ be an accepting run producing (r,m). Choose $k \in \mathbb{N}$ so that $\mathsf{pmin}(\rho) = m - 2k$ or $\mathsf{pmin}(\rho) = m - 2k + 1$ (depending on whether $m - \mathsf{pmin}(\rho)$ is even or odd). Then $0 \ge \mathsf{pmin}(\rho) \ge m - 2k$. Each time operation (ii) is performed from position $\ell \in \mathbb{Z}$, the update is $(r,m) \to (r+3 \cdot 2^{\ell}, m+2)$.

Now, once ρ visits position $\mathsf{pmin}(\rho)$, in order to eventually reach a position $\ell > \mathsf{pmin}(\rho)$, the operation (ii) must be performed on some position $\geq \ell-2$. In particular, to reach position m, it must be performed at some position $m_1 \geq m-2$. If $m_1 > \mathsf{pmin}(\rho)$, to reach m_1 , it must also be performed at some position $m_2 \geq m-4$, etc. Therefore, ρ has to perform (ii) at positions $m_i \geq m-2i$ for each i with $m>m-2i \geq \mathsf{pmin}(\rho)-1$. In other words, it has to do this for each $i=1,\ldots,k$. Each time ρ performs (ii) at m_i , it adds $3 \cdot 2^{m_i}$. Moreover, each extra time ρ performs (ii), it adds a multiple of $3 \cdot 2^{m-2k}$, because $\mathsf{pmin}(\rho) \geq m-2k$. Thus, the number produced in total is some $3n \cdot 2^{m-2k}$ where

$$3n \cdot 2^{m-2k} \ge \sum_{i=1}^{k} 3 \cdot 2^{m_i} \ge \sum_{i=1}^{k} 3 \cdot 2^{m-2i} = f(m,k)$$
.

Conversely, suppose $n \in \mathbb{N}$ and $k \in \mathbb{N}$, $m \in \mathbb{Z}$, $0 \ge m-2k$, and $3n \cdot 2^{m-2k} \ge f(m,k)$. The automaton first moves to position m-2k using operation (i). Then, it performs operations (ii), (i), and (i) again, ℓ times in a loop (we specify ℓ later). That way, it adds $3\ell \cdot 2^{m-2k}$. Then, it moves to position m by applying operation (ii) exactly k times. Hence, it applies (ii) at positions m-2i for $i=1,\ldots,k$ and each time, it adds $3\cdot 2^{m-2i}$. In total, the effect is

$$3\ell \cdot 2^{m-2k} + \sum_{i=1}^{k} 3 \cdot 2^{m-2i} = 3\ell \cdot 2^{m-2k} + f(m,k)$$
.

Since $3n \cdot 2^{m-2k} \ge f(m,k)$ and f(m,k) is an integer multiple of $3 \cdot 2^{m-2k}$, we can choose $\ell \in \mathbb{N}$ so as to produce $3n \cdot 2^{m-2k}$.

Following this example, we first show that any run has the same production as a thin (i.e. bounded thickness) run in which thin returning-left cycles are inserted (p. 10); in the example, such a cycle applies operations (ii), (i), and (i). We then prove that the productions of thin runs form a PE-regular set (p. 12); in the example, the thin run moves to the right to position $\mathsf{pmin}(\rho)$ using operation (i) and then left to $m \ge \mathsf{pmin}(\rho)$ using operations (i) and (ii). Finally, we show that iterating returning-left thin cycles also leads to a PE-regular set (p. 14); in the example, this is how we get all numbers divisible by 3 above a particular bound. We combine these three statements to prove Theorem 3.1.

In combining the thin run with cycles, we will need to ensure that the cycles are anchored on the correct state. To this end, we introduce an annotated version of $\operatorname{pe}([\rho])$ as follows. Let \mathcal{A} be an automaton over $\operatorname{BS}(1,q)$ with state set Q. Let ρ be a run in \mathcal{A} starting and ending in arbitrary states and with $[\rho] = (r,m)$. Letting $\bar{Q} = \{\bar{p} \mid p \in Q\}$ be a copy of Q, we define $\operatorname{sv}(\rho)$, the state view of ρ , to be the word over the alphabet $\Phi_q \cup Q \cup \bar{Q} \cup \{\pm\}$ built as follows. First, write: $\operatorname{pe}([\rho]) = \pm a_{k_1} \cdots a_1 a_0 a_{-1} \cdots a_{-k_2}$, where a_0 has subscript \bullet . Second, let $P_i \in (Q \cup \bar{Q})^{|Q|}$, for $i \in \{-k_2, \ldots, k_1\}$, be a word that contains all the states of Q once in a fixed ordering of Q, either with a bar or not; the states without a bar are exactly those that visit position i in ρ . That is, p appears in P_i iff there is a prefix of ρ ending in p whose final position is i. The state view of ρ is then:

$$\mathrm{sv}(\rho) = \pm a_{k_1} \cdot P_{k_1} \cdots a_0 \cdot P_0 \cdot a_{-1} \cdot P_{-1} \cdots a_{-k_2} \cdot P_{-k_2} \ .$$

We naturally extend sv to sets of runs.

5.1 Any run is equivalent to a thin run augmented with thin returning-left cycles

We now focus on two properties of runs: the states they visit in the automaton and the final position of their prefixes. To that end, we introduce the following notions. For Q a finite set, a position path is a word $\pi \in (Q \times \mathbb{Z})^*$. We extend the analogy with graphs calling elements of $Q \times \mathbb{Z}$ vertices, talking of the vertices visited by a position path, and using the notion of (position) subpaths and cycles. The thickness of a position path π is defined as:

$$\mathsf{thickness}(\pi) = \max_{n \in \mathbb{Z}} |\{i \mid \pi_i = (q,n) \text{ for some } q\}| \enspace .$$

- ▶ **Lemma 5.2.** Let Q be a finite set and $\pi \in (Q \times \mathbb{Z})^*$ be a position path. For any subset V' of the vertices visited by π , there exists a subpath π' of π such that:
- 1. π' starts and ends with the same vertices as π ,
- 2. π' visits all the vertices in V',
- 3. thickness(π') $\leq |Q| \cdot (1 + 2|V'|)$,
- **4.** $\pi \pi'$ consists only of cycles.

Proof. We consider the directed multigraph G that is described by π : the vertices in G are those appearing in π , and an edge appears in G as many times as it does in π . Note that in G, the in- and out-degrees of any vertex are equal, but for the start and end vertices of π .

We first note that Point 4 is true of any subpath π' that satisfies Point 1. Indeed, removing π' from G turns all the vertices into vertices with same in- and out-degrees.

We build π' iteratively. We first let π' be a shortest path from the starting vertex of π to its final vertex in G; since it does not repeat any node in V, its thickness is bounded by |Q|.

Now if π' visits all the vertices in V', we are done. Otherwise, let v be a vertex in V' that π' does not visit; we augment π' with a cycle that includes v as follows. Consider any shortest path from the start vertex of π to v in G, and let u be the last vertex of that path that appears in π' . Write ρ for the path from u to v. Since $\pi - \pi'$ is a union of cycles, there is a path ρ' from v to u in $\pi - \pi'$ (more details follow). We can thus augment π' with the path $\rho\rho'$ rooted at u, potentially increasing the thickness of π' by 2|Q|.

(In more detail, to find the path ρ' , we argue as follows. The set of edges of $\pi - \pi'$ forms an Eulerian multigraph, and so in $\pi - \pi' - \rho$ the difference between outdegree and indegree is 1 for v, -1 for v, and 0 for all other vertices. Therefore, constructing a walk edge by edge, starting from v, while possible, will necessarily lead to a dead end at the vertex u. Removing cycles from this walk will give a path ρ' from v to u, as required.)

▶ Corollary 5.3. Let \mathcal{A} be an automaton over $\mathrm{BS}(1,q)$ with state set Q, and let $k = |Q| + 2|Q|^2$. Any run of \mathcal{A} is equivalent to a run in $\mathrm{Runs}_k(\mathcal{A})$ on which, for each state p appearing in the run, cycles from $\mathrm{RetL}_k^{p\to p}(\mathcal{A})$ are inserted at an occurrence of p with smallest position.

Conversely, any run built by taking a run in $\operatorname{Runs}_k(\mathcal{A})$ and inserting cycles from $\operatorname{RetL}_k^{p\to p}(\mathcal{A})$ at an occurrence of p is a run of \mathcal{A} .

Proof. The converse is clear, we thus focus on the first direction.

(Step 1: Decomposing a run into a thin run and cycles.) Let $\rho \in \mathsf{Runs}(\mathcal{A})$, and extract from it a position path $\pi = \pi_0 \cdots \pi_{|\rho|}$ as follows. We let, $\pi_0 = (q_0, 0)$ and for all $i \geq 1$:

$$\pi_i = (p, n)$$
 where $\rho_i = (\cdot, \cdot, p)$ and $n = \mathsf{pos}(\rho_1 \cdots \rho_i)$.

For each state p visited by ρ , let $n_p = \min\{n \mid \text{there exists } i \text{ such that } \pi_i = (p, n)\}$; in words, n_p is the smallest final position of a prefix of ρ ending in p. Using $V' = \{(p, n_p) \mid \rho \text{ visits } p\}$, Lemma 5.2 provides a position path π' of thickness $\leq k = |Q| + 2|Q|^2$ visiting all of V'.

From π' , we can obtain the corresponding subpath ρ' of ρ that has the same starting and ending state and positions as ρ , and such that ρ is made of ρ' onto which cycles are added. The thickness of ρ' is bounded by k, but the cycles can be of any thickness.

(Step 2: Thinning the cycles.) Consider a cycle β that gets added to ρ' to form ρ , say at position i (after initial i moves, $\rho'_1 \cdots \rho'_i$), and assume that thickness(β) > k. Since a position is repeated more than k > |Q| times, there is a cycle β' within β with thickness(β') $\leq k$; write then $\beta = \alpha \cdot \beta' \cdot \alpha'$. Let p be the state in β' that has the smallest position, that is, p is the ending state of the prefix γ of β' with final position $pmin(\beta')$; write $\beta' = \gamma \cdot \gamma'$. By definition, we have $pos(\rho'_1 \cdots \rho'_i \alpha \gamma) \geq n_p$. Note that $\gamma' \cdot \gamma$ is in $RetL_k^{p \to p}(A)$. We now remove β' from β and then insert $\gamma' \cdot \gamma$ at the position j in ρ' that is such that $\rho'_1 \cdots \rho'_j$ ends in p with final position n_p . For the contribution of $\gamma' \cdot \gamma$ to be the same as that of β' in the original path, we insert it q^d times, where $d = pos(\rho'_1 \cdots \rho'_j \alpha \gamma) - n_p$.

This shows that if any cycle added to ρ' is of thickness > k, then a subcycle of it can be moved to another position of ρ' as a returning-left cycle. Iterating this process, all the cycles added to ρ' will thus be of thickness $\leq k$. Moreover, if an added cycle β is not returning-left after these operations, or if it does not sit at an occurrence of its initial state with *smallest* position, this means that we can decompose it just as above as $\gamma \cdot \gamma'$, with γ reaching $\mathsf{pmin}(\beta)$, and move $\gamma' \cdot \gamma$, a returning-left cycle, to an appropriate position in ρ' as before.

5.2 Intermezzo: reflecting on Corollary 5.3

Before we continue with the proof, we want to illustrate how crucial the previous corollary is. Lemma 5.2 tells us that we can obtain every run from a thin run by then adding cycles. This already simplifies the structure of $\mathsf{Runs}(\mathcal{A})$: indeed, inserting cycles at a certain position in a run $\rho \in \mathsf{Runs}(\mathcal{A})$ corresponds (in algebraic terms) to adding to $[\rho]$ a subset of $\mathbb{Z}[\frac{1}{q}]$ closed under addition, i.e., a submonoid. (Closure under addition follows from the observation that any two returning cycles from each $\mathsf{Ret}_k^{p\to p}(\mathcal{A})$ can be concatenated.)

Sometimes one can conclude that every submonoid of a monoid has a simple structure. For example, every submonoid M of \mathbb{Z} is semilinear and hence a PE-regular subset of $\mathbb{Z}\left[\frac{1}{q}\right]$. Unfortunately, the situation in $\mathbb{Z}\left[\frac{1}{q}\right]$ is not as simple as in \mathbb{Z} :

▶ Fact 5.4. The group $\mathbb{Z}[\frac{1}{a}]$ has uncountably many submonoids.

Proof. Let $q \geq 2$. Consider the functions $f: \mathbb{N} \to \mathbb{Z}$ that satisfy f(0) = 0 and

$$q \cdot f(i) - 1 \le f(i+1) \le q \cdot f(i)$$

for every $i \ge 1$. Note that there are uncountably many such functions f: One can successively choose $f(1), f(2), f(3), \ldots$ and has two options for each value. Consider the set

$$M_f = \left\{ \left. \frac{n}{q^i} \; \right| \; n \ge f(i) \right\}.$$

We claim that for any $n,i\in\mathbb{N}$, we have $\frac{n}{q^i}\in M_f$ iff $n\geq f(i)$. (In other words, it cannot happen that $\frac{n}{q^i}$ can be represented as $\frac{m}{q^j}$ such that $n\geq f(i)$ but not $m\geq f(j)$.) For this, we have to show that $n\geq f(i)$ if and only if $qn\geq f(i+1)$. But if $n\geq f(i)$, then $qn\geq q\cdot f(i)\geq f(i+1)$ by choice of f. Conversely, if $qn\geq f(i+1)$, then $n\geq \frac{1}{q}f(i+1)\geq f(i)-\frac{1}{q}$, which implies $n\geq f(i)$ because n and f(i) are integers. This proves the claim.

The claim implies that M_f is a submonoid of $\mathbb{Z}[\frac{1}{q}]$: For $\frac{n}{q^i}, \frac{m}{q^j} \in M_f$ with $i \leq j$, we have $\frac{n}{q^i} + \frac{m}{q^j} = \frac{q^{j-i}n+m}{q^j}$ and since $m \geq f(j)$, we clearly also have $q^{j-i}n + m \geq f(j)$ and thus $\frac{n}{q^i} + \frac{m}{q^j} \in M_f$. Moreover, since f(0) = 0, we have $0 = \frac{0}{q^0} \in M_f$.

Finally, the claim implies that the mapping $f \mapsto M_f$ is injective: Determining f(i) amounts to finding the smallest $n \in \mathbb{N}$ with $\frac{n}{q^i} \in M_f$.

Thus, $\mathbb{Z}[\frac{1}{q}]$ has submonoids with undecidable membership problem; moreover, there is no hope for a finite description for every submonoid as in \mathbb{Z} . Thus, we need to look at our specific submonoids. A simple observation similar to Lemma 5.2 allows us to obtain every run from a thin part by adding *thin* cycles. Hence, the submonoids that we add are of the form $[\operatorname{Ret}_k^{p\to p}(A)]^*$. It is not hard to show (see Lemma 5.6) that $[\operatorname{Ret}_k^{p\to p}(A)]$ is always a PE-regular set. Thus, one may hope to prove that the regularity of $[\operatorname{Ret}_k^{p\to p}(A)]$ implies regularity of $[\operatorname{Ret}_k^{p\to p}(A)]^*$. (This was an approach to rational subset membership proposed by the third author of this work in [13, Section 4.7].) However, Section 4.3 tells us that even for PE-regular $R \subseteq \mathrm{BS}(1,q)$, the set R^* may not be PE-regular.

Therefore, Corollary 5.3 is the key insight of our proof. It says that a run can be decomposed into a thin part and thin returning-left cycles. Since returning-left cycles produce integers, this will lead us to submonoids of \mathbb{Z} .

5.3 Sets of thin runs are PE-regular

For the proof of that statement, we rely on the following result. It is a classical exercise to show that automata can compute the addition of numbers in a given base. We rely on a slight extension: Using the base-q signed-digit expansion of integers, addition is computable by an automaton:

- ▶ Lemma 5.5 ([8, Section 2.2.2.2]). Let $q \ge 2$ and $B_q = \{-(q-1), \ldots, q-1\}$. Words in B_q^* are interpreted as integers in base q. The language of words over $B_q \times B_q \times B_q$ such that the third component is the sum of the first two components is regular. There is an automaton of size polynomial in q for that language.
- ▶ Lemma 5.6. Let \mathcal{A} be an automaton over $\mathrm{BS}(1,q)$, p,p' be states of \mathcal{A} , and k>0. The sets $\mathrm{sv}(\mathsf{Runs}_k^{p\to p'}(\mathcal{A}))$, $\mathrm{sv}(\mathsf{Ret}_k^{p\to p'}(\mathcal{A}))$, and $\mathrm{sv}(\mathsf{Ret}_k^{p\to p'}(\mathcal{A}))$ are effectively regular.

Proof. For simplicity, we deal with *pointed expansions* of productions of runs, and indicate the easy changes that need to be made to deal with *state views* of runs at the end of the proof. As we draw intuition from two-way automata, we will assume that the positions along a run are always changing. This is easily implemented by changing the alphabet to $\Sigma = \{-1, 0, 1\} \times \{-1, 1\}$, and introducing intermediate states when translating (1, 0) to, say,

(1,1)(0,-1). This modification can turn runs that are k-thin into runs that are 2k-thin: In addition to the k state occurrences from the old run, one also sees at most k state occurrences resulting from non-moving transitions one position to the right. This, however, is not an issue: We perform the construction below for thickness 2k. Then it is obvious from our construction that it can be adapted to only capture those 2k-thin runs in which each original state occurs at most k times in each position.

We will prove the statement in two steps. First, we will convert \mathcal{A} into an automaton that reads k-tuples of letters from $\{-1,0,1\}$. Each component corresponds to one of the "threads" of a run of \mathcal{A} at a given position in the input. Second, we apply Lemma 5.5 to conclude that, based on the regular language over $\{-1,0,1\}^k$ accepted by this new automaton, we can compute the componentwise sum in $\mathbb{Z}\left[\frac{1}{a}\right]$.

(Step 1: From \mathcal{A} to k-component regular language.) This is akin to the classical proof [35] that deterministic two-way automata can be turned into nondeterministic one-way automata. Indeed, since the runs we are interested in are k-thin, we can follow k partial executions of \mathcal{A} , half from left to right, and half from right to left, and check that the reversals of direction are consistent.

In more detail, we will build a nondeterministic automaton \mathcal{B} , whose set of states is $(Q_{\mathcal{A}} \times \{L, R\})^{\leq k}$ and alphabet is $\{-1, 0, 1\}^{\leq k}$. Each component of a given state follows a portion of a k-thin run; it is thus expected that the letters L and R, standing for left and right, and specifying the direction of the partial run, alternate from component to component.

We now specify the transition relation of \mathcal{B} . Let X and Y be two states of \mathcal{B} of the same size $\ell < k$:

$$X = ((p_1, d_1), \dots, (p_\ell, d_\ell)), \quad Y = ((p'_1, d'_1), \dots, (p'_\ell, d'_\ell))$$
.

We add a transition between X and Y labeled (a_1, \ldots, a_ℓ) if for all i:

- $= d_i = d'_i,$
- \blacksquare if $d_i = R$, then $(p_i, (a_i, -1), p'_i)$ is an edge in \mathcal{A} , and
- if $d_i = L$, then $(p'_i, (a_i, 1), p_i)$ is an edge in A.

These transitions check the consistency of a single step. We also add transitions that correspond to the initial and final transitions of runs from p to p' in \mathcal{A} (1 and 2 below), and transitions that check reversals (3 and 4 below):

- 1. At any time, \mathcal{B} can take a transition on ε that either inserts (p, R) as the first component of the current state, or removes (p, L) in that component;
- 2. At any time, \mathcal{B} can take a transition on ε that either inserts (p', L) in the last component of the current state, or removes (p', R) in that component;
- 3. At any time, \mathcal{B} can take a transition on ε that inserts two components (r, L) and (r, R) within the current state, consecutively, for any state r;
- **4.** At any time, \mathcal{B} can take a transition on ε that removes two consecutive components of the form (r, R) and (r, L) from the current state, for any state r.

Naturally, this is subject to the constraint that a state has at most k components. Finally, we set the empty vector as the initial and final state.

To obtain the desired automaton for $\operatorname{\mathsf{Runs}}_k^{p\to p'}(\mathcal{A})$, we additionally modify \mathcal{B} so that transitions of type 1 and 2 are taken exactly once. Moreover, in transition 1, if (p,R) is inserted, then the next symbol read is annotated with \bullet ; if (p,L) is removed, then the previous symbol read is annotated with \bullet . Similarly, transition 2 annotates the next or previous symbol read with \triangleleft .

The automata for $\mathsf{Ret}_k^{p \to p'}(\mathcal{A})$ and $\mathsf{RetL}_k^{p \to p'}(\mathcal{A})$ are obtained by a regular constraint on \mathcal{B} : a simulated run is returning if the symbol annotated with \bullet is also annotated with \triangleleft , and it is returning-left if this is the last symbol.

(Step 2: Computing the addition.) This is a simple application of Lemma 5.5, noting that we can keep the annotations \bullet and \triangleleft as is.

(From pointed expansions to state views.) The automaton \mathcal{B} above actually knows the states in which the different partial runs of \mathcal{A} are; this is what is stored in \mathcal{B} 's states. The alphabet of \mathcal{B} can thus be extended to $(\{-1,0,1\}\times Q)^{\leq k}$, in such a way that each digit carries the information of the state in which it was emitted. Then Step 2 can be changed to not only compute the addition, but also produce the collection of all these states.

5.4 Iterations of returning-left thin cycles are PE-regular

It is well-known that for every set $S \subseteq \mathbb{N}$ the generated monoid $S^* = \{s_1 + \cdots + s_m \mid s_1, \ldots, s_m \in S, m \geq 0\}$ is eventually identical with $\gcd(S) \cdot \mathbb{N}$. In other words, the set $(\gcd(S) \cdot \mathbb{N}) \setminus S^*$ is finite and we may define $F(S) = \max((\gcd(S) \cdot \mathbb{N}) \setminus S^*)$. The number F(S) is called the *Frobenius number* of S. With this, we have $S^* = \{n \in S^* \mid n \leq F(S)\} \cup \{n \in \gcd(S) \cdot \mathbb{N} \mid n > F(S)\}$. If $S \subseteq -\mathbb{N}$, then we set F(S) := F(-S). Now consider an arbitrary set $S \subseteq \mathbb{Z}$. If S contains both a positive and a negative number, then $S^* = \gcd(S) \cdot \mathbb{Z}$ and we set F(S) := 0. We shall use the following well-known fact [39].

- ▶ **Lemma 5.7.** If $S = \{n_1, ..., n_k\}$ with $0 < n_1 < ... < n_k$, then $F(S) \le n_k^2$.
- ▶ Lemma 5.8. For every automaton \mathcal{A} over $\mathrm{BS}(1,q)$, the language $\mathrm{pe}([\mathrm{RetL}_k^{p\to p}(\mathcal{A})]^*)$ is effectively regular.

Proof. Recall that we identify each $r \in \mathbb{Z}[\frac{1}{q}]$ with $(r,0) \in \mathbb{Z}[\frac{1}{q}]$. In particular, for $n \in \mathbb{Z}$, pe(n) is the same as pe((n,0)).

Denote $S = [\mathsf{RetL}_k^{p \to p}(\mathcal{A})]$. We first consider the case $S \subseteq \mathbb{N}$ and $S \neq \emptyset$. Suppose we can compute $\gcd(S)$ and a bound $B \in \mathbb{N}$ with $B \geq F(S)$. Then we have

$$S^* = \underbrace{\{n \in S^* \mid n \le B\}}_{=:X} \cup \underbrace{\{n \in \gcd(S) \cdot \mathbb{N} \mid n > B\}}_{=:Y}$$
 (2)

and it suffices to show that $\operatorname{pe}(X)$ and $\operatorname{pe}(Y)$ are effectively regular. Note that X is finite and can be computed by finding all $n \leq B$ with $n \in S$ (recall that membership in S is decidable because $\operatorname{sv}(\operatorname{RetL}_k^{p \to p}(\mathcal{A}))$ is effectively regular by Lemma 5.6) and building sums. Moreover, $\operatorname{pe}(Y)$ is regular because the set $L_0 = \operatorname{pe}(\gcd(S) \cdot \mathbb{N})$ is effectively regular and so is $L_1 = \{\operatorname{pe}(n) \mid n \in \mathbb{N}, n > B\}$, and hence $\operatorname{pe}(Y) = L_0 \cap L_1$.

Thus, it remains to compute $\gcd(S)$ and some $B \geq F(S)$. For the former, find any $r \in S$ and consider its decomposition $r = p_1^{e_1} \cdots p_m^{e_m}$ into prime powers. For each $i \in [1, m]$, we compute $d_i \in [0, e_i]$ and $n_i \in S$ such that (i) $S \subseteq p_i^{d_i} \cdot \mathbb{N}$, and (ii) $n_i \in S \setminus p_i^{d_i+1} \cdot \mathbb{N}$. Since for $d \in \mathbb{N}$, we can construct an automaton for $\operatorname{pe}(S \cap d \cdot \mathbb{N})$, these d_i and n_i can be computed. Observe that $\gcd(S) = p_1^{d_1} \cdots p_m^{d_m}$. Let $T = \{r, n_1, \dots, n_k\}$. Observe that $\gcd(T) = \gcd(S)$, and hence T^* and S^* are ultimately identical. Since $T \subseteq S$, this means $F(S) \leq F(T)$. By Lemma 5.7, we have $F(T) \leq (\max\{r, n_1, \dots, n_k\})^2$, which yields our bound B.

The case $S \subseteq -\mathbb{N}$ is analogous to $S \subseteq \mathbb{N}$. If S contains a positive and a negative number, then $S^* = \gcd(S) \cdot \mathbb{Z}$, so it suffices to just compute $\gcd(S)$. This is done as above. Finally, deciding between these three cases is easy. This completes the proof.

5.5 Wrapping up: Proof of Theorem 3.1

Let \mathcal{A} be an automaton over BS(1,q) with state set Q. Corollary 5.3 indicates that the set of productions of accepting runs is the same as the set of productions of k-thin runs in which thin cycles are introduced.

By Lemma 5.6, $\operatorname{sv}(\operatorname{Runs}_k(\mathcal{A}))$ is a regular language L. For any state p of \mathcal{A} , let $L_p = \operatorname{pe}([\operatorname{RetL}_k^{p\to p}(\mathcal{A})]^*)$, a regular language by Lemma 5.8. For padding purposes, let $s\in Q$ be some state, and let h be the morphism from $(\Phi_q \cup \{\pm\})^*$ to $(\Phi_q \cup Q \cup \{\pm\})^*$ defined, for any $a\in \Phi_q$, by $h(a)=as^{|Q|}$, and h(+)=+, h(-)=-. Define now L_p' to be the image by h of the version of L_p where arbitrary 0's are added after the sign, and at the end of the number (these 0's do not change the value represented).

Consider now the language R over the alphabet $(\Phi_q \cup Q \cup \bar{Q} \cup \{\pm\})^{|Q|+1}$ whose projection on the first component is the language L, and the other components correspond to the languages L'_p , for each $p \in Q$. The first component indicates in particular the states of \mathcal{A} that visited that location; to synchronize the different components of R, we ensure that the letter annotated with \bullet in L'_p is aligned with a letter from L that is followed by p—that is, the starting position of L'_p is at a position in L that is seen while being in the state p.

Finally, an automaton can do the componentwise addition in base q, collapsing the |Q|+1 components into a single one. The radix point is given by the digit with \bullet of L, i.e., in the first component; and similarly for \triangleleft . The resulting language, thanks to Corollary 5.3, is the language of the pointed expansions of all runs in $\mathsf{Runs}(A)$.

6 Complexity

In this section, we prove Theorems 3.2 and 3.3. For the upper bounds in Theorems 3.2 and 3.3, we shall rely on the fact that, given an element $g \in \mathbb{Z}[\frac{1}{q}] \rtimes \mathbb{Z}$ as a word over $\Sigma = \{a, a^{-1}, t, t^{-1}\}$, one can compute the pointed expansion $\operatorname{pe}(g)$ in logarithmic space. This is a direct consequence of a result of Elder, Elston, and Ostheimer [19, Proposition 32]. They show that given a word w over Σ , one can compute in logarithmic space an equivalent word of one of the forms (i) t^i , (ii) $(a^{\eta_0})^{t^{\alpha_0}}(a^{\eta_1})^{t^{\alpha_1}}\cdots(a^{\eta_k})^{t^{\alpha_k}}t^i$ or (iii) $(a^{-\eta_0})^{t^{\alpha_0}}(a^{-\eta_1})^{t^{\alpha_1}}\cdots(a^{-\eta_k})^{t^{\alpha_k}}t^i$, where $i \in \mathbb{Z}$, $k \in \mathbb{N}$, $0 < \eta_j < q$ for $j \in [0, k]$, and $\alpha_0 > \cdots > \alpha_k$. Here, x^y stands for $y^{-1}xy$ in the group. Since these normal forms denote the elements (i) (0, i), (ii) $(\sum_{j=0}^k \eta_j q^{-\alpha_j}, i)$ and (iii) $(-\sum_{j=0}^k \eta_j q^{-\alpha_j}, i)$, respectively, it is easy to turn these normal forms into $\operatorname{pe}(w)$ using logarithmic space.

This allows us to prove Theorem 3.3: For every rational subset $R \subseteq BS(1,q)$, the language pe(R) is a regular language. In particular, there exists a deterministic automaton \mathcal{B} for pe(R). Therefore, given $g \in BS(1,q)$ as a word over $\{a,a^{-1},t,t^{-1}\}$, we compute pe(g) in logspace and then check membership of pe(g) in $L(\mathcal{B})$, which is decidable in logarithmic space.

6.1 PSPACE-hardness

The PSPACE lower bound in Theorem 3.2 is a reduction from the intersection nonemptiness of finite-state automata, a well-known PSPACE-complete problem [25].

▶ **Theorem 6.1.** *Rational subset membership is* PSPACE-*hard.*

Proof. Let $q \geq 2$ be fixed. We give a reduction from the intersection nonemptiness problem for deterministic finite automata (DFA), a PSPACE-hard problem [25]. Let $\mathcal{D}_1, \ldots, \mathcal{D}_n$, DFA over a finite alphabet Γ , $|\Gamma| \geq 2$, form an instance of that problem. We will describe an automaton \mathcal{A} over BS(1, q) that accepts the identity element of BS(1, q) if and only if there is a word $w \in \Gamma^*$ accepted by all \mathcal{D}_i .

We first fix any injective mapping $f \colon \Gamma \to \{0, 1, \dots, q-1\}^{\ell}$ for $\ell = \lceil \log_2 |\Gamma| \rceil$. Transform $\mathcal{D}_1, \dots, \mathcal{D}_n$ into nondeterministic finite automata (NFA) $\mathcal{D}'_1, \dots, \mathcal{D}'_n$ over $\{0, 1, \dots, q-1\}$ such that $\mathsf{L}(\mathcal{D}'_i) = 1 \cdot f(\mathsf{L}(\mathcal{D}_i)) \cdot 1$ for all i. It is immediate that $\mathsf{L}(\mathcal{D}_1) \cap \dots \cap \mathsf{L}(\mathcal{D}_n)$ is nonempty if and only if so is $\mathsf{L}(\mathcal{D}'_1) \cap \dots \cap \mathsf{L}(\mathcal{D}'_n)$.

We now describe the construction of the automaton \mathcal{A} ; it will be convenient for us to think of the input word as being written (produced) rather than read by \mathcal{A} . This word over $\{-1,0,1\} \times \{-1,1\} \subseteq \mathrm{BS}(1,q)$ corresponds to instructions to a machine working over an infinite tape with alphabet $\{0,1,\ldots,q-1\}$, as per the intuition explained in Section 2, and we will think of \mathcal{A} as moving left and right over that tape, updating the values in its cells. We emphasize that this tape is *not* the input tape of \mathcal{A} , but instead corresponds to the actions of generators of $\mathrm{BS}(1,q)$.

The automaton \mathcal{A} will subdivide the tape into n tracks. Suppose the cells of the tape are numbered, with indices $m \in \mathbb{Z}$; then the ith track consists of all cells with indices x such that $x \equiv i \mod n$. The automaton \mathcal{A} will move left and right over the tape by producing t = (0,1) and $t^{-1} = (0,-1)$, two of the generators of $\mathrm{BS}(1,q)$ as monoid. Similarly, the current cell can be updated by producing a = (1,0) and $a^{-1} = (-1,0)$, i.e., performing increments and decrements. The automaton will always remember in its finite-state memory which of the tracks the current cell belongs to.

The workings of \mathcal{A} are as follows. It will enumerate $i=1,\ldots,n$ one by one, and for each i it will guess and print some word accepted by the NFA \mathcal{D}'_i on the ith track of the tape. (When we refer to guessing, this corresponds to the nondeterminism in the definition of automata over groups.) When incrementing i, it will not only move to the (i+1)st track but also guess which specific cell in this track to move to. That is, in principle, \mathcal{A} may move arbitrarily far left or right over the tape. After all values of i have been enumerated, the automaton \mathcal{A} will guess some position of track 1 on the tape, moving to that position. Suppose the corresponding cell is numbered $x \in \mathbb{Z}$, $x \equiv 1 \mod n$; then \mathcal{A} will transition to its final phase, performing the following sequence of operations:

1. For i = 1, ..., n: perform decrement of the cell value once (k = 1 times), and then move to the adjacent cell with larger index (thus proceeding to track i + 1, or to track 1 again if i = n).

We think of this sequence of instructions as the removal of k, k = 1.

- 2. Perform the following operations in a loop, taken arbitrarily many times (terminating after some nondeterministically chosen iteration):
 - Guess an element $g \in \{0, 1, \dots, q-1\}$.
 - \blacksquare Remove g (similarly to step 1).
- **3.** $Remove\ 1$ (as in step 1).
- **4.** Move to an arbitrarily chosen cell of the tape and terminate (i.e., transition to a final state).

We now claim that the final configuration of the tape can be all-0 (i.e., the produced generators of BS(1,q) can yield the identity element of BS(1,q)) if and only if there is a word accepted by all machines \mathcal{D}'_i , $i=1,\ldots,n$.

Indeed, observe that, by the construction of \mathcal{A} , at the end of the simulation of NFA $\mathcal{D}'_1, \ldots, \mathcal{D}'_n$ each track i will contain a word of the form $1 \cdot f(w_i) \cdot 1$ where $w_i \in \mathsf{L}(\mathcal{D}_i)$, with zeros all around it. The words written on different tracks may or may not be aligned with each other. Clearly, if all w_i are chosen to be the same word, w, and the leftmost 1s are all aligned with each other, then in the final phase of computation the automaton \mathcal{A} can guess the word w and v and v are completely (with delimiters). After that, it can guess the location of cell 0 and move to that cell—this corresponds to the product of the produced generators being the identity of $\mathsf{BS}(1,q)$.

Therefore, it remains to see that the final phase cannot transform the tape configuration to all-0 unless all words w_i are the same and the delimiting 1s are aligned. But for this, it suffices to observe that the final phase (excepting the last operation) amounts, in terms of

the group BS(1,q), to subtracting a number of the following form (written in base q):

$$\underbrace{1\ldots 1}_{n}\underbrace{g_{1}\ldots g_{1}}_{n}\cdots\underbrace{g_{s}\ldots g_{s}}_{n}\underbrace{1\ldots 1}_{n},$$

where $s \in \mathbb{N}$ and $g_1, \ldots, g_s \in \{0, 1, \ldots, q-1\}$ are chosen nondeterministically by \mathcal{A} . If the result of subtraction is $0 \in \mathbb{Z}$, then the content of the tape did indeed correspond to a number of this form. So the simulation of phase left each track with the same content, $\ldots 001g_1 \ldots g_s 100 \ldots$, which means that $f^{-1}(g_1 \ldots g_s) \in \mathsf{L}(\mathcal{D}_1) \cap \ldots \cap \mathsf{L}(\mathcal{D}_n)$.

Since the construction of the automaton \mathcal{A} can be performed in polynomial time (and even in logarithmic space), this completes the proof.

6.2 **PSPACE** membership

For the PSPACE upper bound, we strengthen Theorem 3.1 by constructing a polynomial-size representation of an exponential size automaton for the resulting regular language. A succinct finite automaton is a tuple $\mathcal{S} = (n, \Gamma, (\varphi_x)_{x \in \Gamma \cup \{\varepsilon\}}, p_0, p_f\})$, where $n \in \mathbb{N}$ is its bit length, Γ is its input alphabet, $\varphi_x(\mathsf{v}_1, \ldots, \mathsf{v}_n, \mathsf{v}'_1, \ldots, \mathsf{v}'_n)$ is a formula from propositional logic with free variables $\mathsf{v}_1, \ldots, \mathsf{v}_n, \mathsf{v}'_1, \ldots, \mathsf{v}'_n$ for each $x \in \Gamma \cup \{\varepsilon\}$, $p_0 \in \{0, 1\}^n$ is its initial state, and $p_f \in \{0, 1\}^n$ is its final state. The size of \mathcal{S} is defined as $|\mathcal{S}| = n + \sum_{x \in \Gamma \cup \{\varepsilon\}} |\varphi_x|$, where $|\varphi|$ denotes the length of the formula φ .

Moreover, S represents the automaton A(S), which is defined as follows. It has the state set $\{0,1\}^n$, initial state p_0 , and final state p_f . For states $p=(b_1,\ldots,b_n), p'=(b'_1,\ldots,b'_n) \in \{0,1\}^n$ and $x \in \Gamma \cup \{\varepsilon\}$, there is an edge (p,x,q) in A(S) if and only if $\varphi_x(b_1,\ldots,b_n,b'_1,\ldots,b'_n)$ holds. We define the language accepted by S as L(S) = L(A(S)).

We allow ε -edges in succinct automata, and with Boolean formulas, one can encode steps in a Turing machine. Thus, a succinct automaton of polynomial size can simulate a polynomial space Turing machine with a one-way read-only input tape. Our descriptions of succinct automata will therefore be in the style of polynomial space algorithms. We show:

▶ Theorem 6.2. Given a rational subset $R \subseteq BS(1,q)$, one can construct in polynomial space a polynomial-size succinct automaton accepting pe(R).

This allows us to decide rational subset membership in PSPACE: Given an automaton \mathcal{A} over BS(1,q) and an element g as a word over $\{a,a^{-1},t,t^{-1}\}$, we construct a succinct automaton \mathcal{B} for $pe(L(\mathcal{A}))$ and the pointed expansion pe(g) in logarithmic space. Since membership in succinct automata is well-known to be in PSPACE, we can check whether $pe(g) \in L(\mathcal{B})$.

Constructing succinct automata It remains to prove Theorem 6.2. The construction of a succinct automaton for pe(R) proceeds with the same steps as in Section 5. For most of these steps, our constructions already yield small succinct automata (e.g., one for $pe([RetL_k^{p\to p'}(A)])$ in Lemma 5.6). The exception is Lemma 5.8 — in which case the key ingredient is as follows.

▶ Proposition 6.3. Given an automaton \mathcal{A} over $\mathrm{BS}(1,q)$, a state p of \mathcal{A} , and $k \in \mathbb{N}$ in unary, one can compute in polynomial space the number $\gcd([\mathsf{RetL}_k^{p\to p}(\mathcal{A})])$ and a bound $B \geq F([\mathsf{RetL}_k^{p\to p}(\mathcal{A})])$. Both are at most exponential in k and the size of \mathcal{A} .

Our bound on F extends the bound for automatic sets in \mathbb{N} [5, Lemma 4.5] to thin two-way computations. Before proving Proposition 6.3, let us show how it implies Theorem 6.2.

Proof of Theorem 6.2. The constructions in Lemma 5.6 and Theorem 3.1, immediately yield a polynomial-size succinct automaton for pe(R) once a succinct automaton for each

pe([RetL $_k^{p\to p}(A)$]*) is found. For the latter, we proceed as in Lemma 5.8. Let $S=[\text{RetL}_k^{p\to p}(A)]$ and compute $\gcd(S)$ and a bound $B\geq F(S)$ using Proposition 6.3. Then, by Equation (2) on page 14, it suffices to construct a succinct automaton for $\operatorname{pe}(X)$ and one for $\operatorname{pe}(Y)$. For $\operatorname{pe}(X)$, we use the fact that we can construct a succinct automaton $\mathcal B$ for $\operatorname{pe}(S)$. Our automaton for $\operatorname{pe}(X)$ proceeds as follows. With ε -transitions, it runs $\mathcal B$ to successively guess numbers $\leq B$ from S and stores each of them temporarily in its state. Such a number requires $O(\log(B))$ bits. In another $O(\log(B))$ bits, it stores the sum of the numbers guessed so far. This continues as long as the sum is at most B. Then, our automaton reads the resulting sum from the input. This automaton clearly accepts $\operatorname{pe}(X)$.

For pe(Y), we have to construct a succinct automaton that accepts any number > B that is divisible by gcd(S). Since gcd(S) is available as a number with polynomially many digits, we can construct a succinct automaton accepting $pe(gcd(S) \cdot \mathbb{N})$: It keeps the remainder modulo gcd(S) of the currently read prefix. This requires $O(\log(gcd(S)))$ many bits. Since B also has polynomially many digits, we can construct a succinct automaton for $\{n \in \mathbb{N} \mid n > B\}$. An automaton for the intersection then accepts pe(Y).

It is easy to see that the number produced by a returning-left run is at most exponential in the length of the run. The exact bound will not be important.

▶ **Lemma 6.4.** If ρ is a run in $\operatorname{RetL}_k(\mathcal{A})$ of length ℓ , then $|[\rho]| \leq q^{2\ell}$.

Proof. Let $m = \mathsf{pmax}(\rho)$. Since ρ is returning-left, m can be at most $\ell/2$. Suppose in each position $i \in [0, m]$, ρ adds $x_i \cdot q^i$. Then we have $|x_0| + \cdots + |x_m| \le \ell - 2m$ and also

$$|[\rho]| = |x_0q^0 + \dots + |x_mq^m| \le |x_0|q^0 + \dots + |x_m|q^m.$$

Under the condition $|x_0|+\cdots+|x_m|\leq \ell-2m$, the expression on the right is clearly maximized for $x_m=\ell-2m$ and $x_i=0$ for $i\in[0,m-1]$. Therefore, we have $|[\rho]|\leq (\ell-2m)q^m$. Since $\ell-2m\leq q^\ell$, this implies $|[\rho]|\leq (\ell-2m)q^m\leq q^\ell\cdot q^{\ell/2}\leq q^{2\ell}$.

The main ingredient for Proposition 6.3 will be Lemma 6.5. We write $\rho \ll \rho'$ if $|\rho| < |\rho'|$. Moreover, for $d \in \mathbb{Z}$, we write $\rho \ll_d \rho'$ if $\rho \ll \rho'$ and for some $\ell \in \mathbb{Z}$, we have $[\rho'] = \ell \cdot [\rho] + d$.

▶ **Lemma 6.5.** There is a polynomial f such that the following holds. Let \mathcal{A} be an n-state automaton over $\mathrm{BS}(1,q)$ and let p,p' be two states of \mathcal{A} . Let $\rho_{11} \in \mathrm{RetL}_k^{p \to p'}(\mathcal{A})$ with $|\rho_{11}| > f(n,k)$. There exist runs $\rho_{00}, \rho_{10}, \rho_{01} \in \mathrm{RetL}_k^{p \to p'}(\mathcal{A})$ and $d \in \mathbb{Z}$ so that:

$$\begin{array}{ccc}
\rho_{01} & \ll_d & \rho_{11} \\
 & & & & & \\
\rho_{00} & \ll_d & \rho_{10}
\end{array} \tag{3}$$

Here, one shows that a long run can be shortened independently in two ways: Going left in the diagram (3), and going down. Shortening the run by "going left" changes the production of the run by the same difference, up to a factor ℓ that may differ in the two rows.

In order to prove Lemma 6.5, we first show a version of Lemma 6.5 that applies to returning runs that go far to the left and far to the right. We first show some auxiliary lemmas:

▶ Lemma 6.6. Let \mathcal{A} be an n-state automaton over $\mathrm{BS}(1,q)$. For every $\tau \in \mathrm{Ret}_k(\mathcal{A})$ with $\mathrm{pmax}(\tau) > n^2$ or $\mathrm{pmin}(\tau) < -n^2$, there is a run $\tau' \in \mathrm{Ret}_k(\mathcal{A})$ with $|\tau'| < |\tau|$ and $\mathrm{pmin}(\tau') \geq \mathrm{pmin}(\tau)$. Moreover, τ' begins and ends in the same states as τ .

Proof. If we consider the effect of the actions of \mathcal{A} on the cursor, then the statement amounts to the following statement on n-state one-counter automata with a \mathbb{Z} -counter and a single zero test at the end of (accepted) runs: if along a run τ the counter goes (i) above n^2 or (ii) below $-n^2$, then there is a strictly shorter run τ' which begins and ends in the same states and in which the minimum value of the counter is at least the minimum value of the counter in τ . This can be proved using the standard hill-cutting argument (see, e.g., [20, Lemma 5]; cf. [26, Proposition 7] as well as [9] and references therein): in scenario (i) one can apply it to reduce the *maximum* value of the counter whilst retaining the minimum value; and in scenario (ii) one can increase the *minimum* value whilst retaining the maximum one.

The following is a consequence of Lemma 6.6.

- ▶ **Lemma 6.7.** There is a polynomial f so that for every n-state automaton \mathcal{A} over BS(1,q) and every two states p, p' of \mathcal{A} , the shortest run in $RetL_{k}^{p \to p'}(\mathcal{A})$ has length $\leq f(n,k)$.
- ▶ **Lemma 6.8.** If ρ is a run in RetL_k(\mathcal{A}), then $|\rho|/k < \text{pmax}(\rho) + 1$.

Proof. The run ρ can visit at most $\operatorname{pmax}(\rho)+1$ distinct positions. But since ρ is k-thin, it can visit each position at most k times. Since ρ has $|\rho|$ moves, we have $|\rho|+1 \leq k(\operatorname{pmax}(\rho)+1)$ and thus $|\rho| < k(\operatorname{pmax}(\rho)+1)$, hence $|\rho|/k < \operatorname{pmax}(\rho)+1$.

We now turn to our simpler version of Lemma 6.5. For runs ρ, ρ' and $d \in \mathbb{Z}$, we write $\rho \ll_{(d)} \rho'$ if $|\rho| < |\rho'|$ and $[\rho'] = [\rho] + d$.

▶ **Lemma 6.9.** Let \mathcal{A} be an n-state automaton over $\mathrm{BS}(1,q)$ and let p,p' be two states of \mathcal{A} . Suppose $\rho_{11} \in \mathrm{Ret}_k^{p \to p'}(\mathcal{A})$ is such that $\mathrm{pmax}(\rho_{11}) > n^2$ and $\mathrm{pmin}(\rho_{11}) < -n^2$. Then there are runs $\rho_{00}, \rho_{10}, \rho_{01}, \rho_{11} \in \mathrm{Ret}_k^{p \to p'}(\mathcal{A})$ and a number $d \in \mathbb{Z}$ so that $\mathrm{pmin}(\rho) \geq \mathrm{pmin}(\rho_{11})$ for $\rho \in \{\rho_{00}, \rho_{01}, \rho_{10}\}$ and the following holds:

$$\begin{array}{ccc}
\rho_{01} & \ll_{(d)} & \rho_{11} \\
 & & & & & \\
\rho_{00} & \ll_{(d)} & \rho_{10}
\end{array} \tag{4}$$

Proof. Since $\operatorname{pmax}(\rho_{11}) > n^2$ and $\operatorname{pmin}(\rho_{11}) < -n^2$, we can decompose $\rho_{11} = \sigma_1 \tau_1 \nu_1$ such that $\sigma_1, \nu_1 \in \operatorname{Ret}_k(\mathcal{A})$ and $\tau_1 \in \operatorname{RetL}_k(\mathcal{A})$ and $\operatorname{pmax}(\tau_1) > n^2$ and either $\operatorname{pmin}(\sigma_1) < -n^2$ or $\operatorname{pmin}(\nu_1) < -n^2$. (Note that none of σ_1 , τ_1 , ν_1 needs to be a cycle.) Without loss of generality, we assume $\operatorname{pmin}(\nu_1) < -n^2$.

According to Lemma 6.6, there are $\nu_0, \tau_0 \in \mathsf{Ret}_k(\mathcal{A})$ with $|\nu_0| < |\nu_1|$ and $|\tau_0| < |\tau_1|$ and $\mathsf{pmin}(\nu_0) \ge \mathsf{pmin}(\nu_1)$ and $\mathsf{pmin}(\tau_0) \ge \mathsf{pmin}(\tau_1)$. Since $\tau_1 \in \mathsf{RetL}_k(\mathcal{A})$, this implies $\tau_0 \in \mathsf{RetL}_k(\mathcal{A})$. Define

$$\rho_{01} = \sigma_1 \tau_0 \nu_1 \qquad \qquad \rho_{11} = \sigma_1 \tau_1 \nu_1
\rho_{00} = \sigma_1 \tau_0 \nu_0 \qquad \qquad \rho_{10} = \sigma_1 \tau_1 \nu_0$$

Then with $d = [\tau_1] - [\tau_0]$, we have $[\rho_{1i}] = [\rho_{0i}] + d$ for i = 0 and i = 1.

We are now prepared to prove Lemma 6.5.

Proof of Lemma 6.5. Write $\rho = \rho_{11}$, let $f(n,k) = 3k(n^2+1)$, and suppose $|\rho| > 3k(n^2+1)$. Then $\mathsf{pmax}(\rho) + 1 > 3(n^2+1)$ by Lemma 6.8, so $\mathsf{pmax}(\rho) \ge 3(n^2+1)$ and, in particular, we can decompose $\rho = \sigma \tau \nu$ so that σ is the shortest prefix of ρ with $\mathsf{pmax}(\sigma) = 2(n^2+1)$ and $\sigma \tau$ is the longest prefix of ρ with $\mathsf{pmax}(\sigma \tau) = 2(n^2+1)$. Since $\mathsf{pmax}(\rho) \ge 3(n^2+1)$, we have $\mathsf{pmax}(\tau) > n^2$. We distinguish two cases.

$$\begin{array}{ccc} \tau_{01} & \ll_{(d)} & \tau_{11} = \tau \\ & & & & \\ \tau_{00} & \ll_{(d)} & \tau_{10} \end{array}$$

and $pmin(\tau') \ge pmin(\tau)$ for every $\tau' \in \{\tau_{00}, \tau_{01}, \tau_{10}\}$. We set $\rho_{ij} = \sigma \tau_{ij} \nu$. Then each ρ_{ij} belongs to $RetL_k^{p \to p'}(\mathcal{A})$ and we even have

$$\begin{array}{ccc} \rho_{01} & \ll_{(d)} & \rho_{11} \\ & & & & \forall \\ \rho_{00} & \ll_{(d)} & \rho_{10} \end{array}$$

which implies Equation (3).

2. Suppose $pmin(\tau) \ge -n^2$. In this case, Lemma 6.6 yields a run $\tau' \in Ret_k(\mathcal{A})$ with $|\tau'| < |\tau|$ and $pmin(\tau') \ge pmin(\tau)$.

Since now $pmin(\tau) \ge -n^2$ and $pmin(\tau') \ge -n^2$, we can decompose $\sigma = \sigma_1 \sigma_2 \sigma_3$ and $\nu = \nu_3 \nu_2 \nu_1$ so that

- $|\sigma_2| > 0$ and $|\nu_2| > 0$ and
- $= pos(\sigma_1) + pos(\nu_1) = 0 \text{ and } pos(\sigma_2) + pos(\nu_2) = 0.$
- $\sigma_1 \sigma_3 \tau \nu_3 \nu_1$ and $\sigma_1 \sigma_3 \tau' \nu_3 \nu_1$ again belong to $\mathsf{RetL}_k(\mathcal{A})$.

For ease of notation, we write $\tau_1 = \tau$ and $\tau_0 = \tau'$. We define

$$\rho_{01} = \sigma_1 \sigma_3 \tau_1 \nu_3 \nu_1$$

$$\rho_{11} = \sigma_1 \sigma_2 \sigma_3 \tau_1 \nu_3 \nu_2 \nu_1$$

$$\rho_{00} = \sigma_1 \sigma_3 \tau_0 \nu_3 \nu_1$$

$$\rho_{10} = \sigma_1 \sigma_2 \sigma_3 \tau_0 \nu_3 \nu_2 \nu_1$$

(where ρ_{11} is repeated just for illustration). Then clearly the length relationships claimed in Equation (3) are satisfied. Let $h_1 = pos(\sigma_1)$ and $h_2 = pos(\sigma_2)$. Then both for i = 0 and for i = 1, we have

$$\begin{split} [\rho_{1i}] &= [\sigma_1] + q^{h_1}[\sigma_2] + q^{h_1 + h_2}[\sigma_3 \tau_i \nu_3] + q^{h_1 + h_2}[\nu_2] + q^{h_1}[\nu_1] \\ [\rho_{0i}] &= [\sigma_1] + q^{h_1}[\sigma_3 \tau_i \nu_3] + q^{h_1}[\nu_1]. \end{split}$$

Therefore, with $d = (1 - q^{h_2})[\sigma_1] + q^{h_1}[\sigma_2] + q^{h_1 + h_2}[\nu_2] + q^{h_1}(1 - q^{h_2})[\nu_1]$, we have

$$\begin{aligned} [\rho_{1i}] &= q^{h_2}[\rho_{0i}] + [\sigma_1] - q^{h_2}[\sigma_1] + q^{h_1}[\sigma_2] + q^{h_1+h_2}[\nu_2] + q^{h_1}[\nu_1] - q^{h_1+h_2}[\nu_1] \\ &= q^{h_2}[\rho_{0i}] + d \end{aligned}$$

This means that indeed $\rho_{01} \ll_d \rho_{11}$ and $\rho_{00} \ll_d \rho_{10}$.

Lemma 6.10 applies Lemma 6.5 to construct small numbers in $[\mathsf{RetL}_k(\mathcal{A})]$ that are not divisible by a given m. Later, these numbers allow us to compute $\gcd([\mathsf{RetL}_k^{p\to p}(\mathcal{A})])$ and bound $F([\mathsf{RetL}_k^{p\to p}(\mathcal{A})])$.

▶ Lemma 6.10. There is a polynomial f such that the following holds. Let $m \in \mathbb{Z}$. Let \mathcal{A} be an n-state automaton over $\mathrm{BS}(1,q)$ and let p,p' be two states of \mathcal{A} . Suppose there is a number in $[\mathsf{RetL}_k^{p \to p'}(\mathcal{A})]$ not divisible by m; then there is also an $s \in [\mathsf{RetL}_k^{p \to p'}(\mathcal{A})]$ not divisible by m such that $|s| \leq q^{f(n,k)}$.

Proof. Let f be the polynomial from Lemma 6.5. Let $\rho \in \text{RetL}_k^{p \to p'}(\mathcal{A})$ be of minimal length such that m does not divide $[\rho]$. Suppose $|\rho| > f(n,k)$. Write $\rho_{11} = \rho$ and apply Lemma 6.5. By minimality of ρ_{11} , we get $[\rho_{00}] \equiv [\rho_{10}] \equiv [\rho_{01}] \equiv 0 \mod m$. In particular, $\rho_{00} \ll_d \rho_{10}$ implies $d \equiv 0 \mod m$. However, since $\rho_{01} \ll_d \rho_{11}$ and $[\rho_{11}] \not\equiv 0 \mod m$, we get $d \not\equiv 0 \mod m$, a contradiction. Hence, $|\rho| \leq f(n,k)$ and thus $|[\rho]| \leq q^{2f(n,k)}$ by Lemma 6.4.

With Lemma 6.10 in hand, one can show Proposition 6.3 similarly to Lemma 5.8.

Proof of Proposition 6.3. Denote $S = [\text{RetL}_k^{p \to p}(\mathcal{A})]$ and suppose $S \neq \emptyset$. Let f_1 be the polynomial from Lemma 6.7. Then the shortest run in $\text{RetL}_k^{p \to p}(\mathcal{A})$ has length $\leq f_1(n,k)$. We can therefore guess a run ρ of length $\leq f_1(n,k)$.

If we write $r = [\rho]$, then $|r| \leq q^{2f_1(n,k)}$ by Lemma 6.4. We can thus compute r in polynomial space. Note that $g = \gcd(S)$ divides r and thus $g \leq q^{2f_1(n,k)}$. Let us now describe how to compute g and a bound $B \geq F(S)$.

We first consider the case $S \subseteq \mathbb{N}$. We compute the decomposition $r = p_1^{e_1} \cdots p_m^{e_m}$ into prime powers. Note that each e_i is at most polynomial. For each $i \in [1, m]$, there exists a $d_i \in [0, e_i]$ such that $S \subseteq p_i^{d_i} \cdot \mathbb{N}$ but $S \not\subseteq p_i^{d_i+1} \cdot \mathbb{N}$. We can compute d_i in polynomial space, because we can construct a succinct finite automaton for $\operatorname{pe}(S)$ and, for every polynomially bounded ℓ , we can construct a succinct automaton for $\operatorname{pe}(\mathbb{N} \setminus p_i^{\ell} \cdot \mathbb{N})$: The latter keeps a remainder modulo p_i^{ℓ} in its state, accepting if this remainder is non-zero. Thus, given a candidate d_i , we can construct a succinct automaton for $\operatorname{pe}(S \cap (\mathbb{N} \setminus p_i^{d_i} \cdot \mathbb{N}))$ and one for $\operatorname{pe}(S \cap (\mathbb{N} \setminus p_i^{d_i+1} \cdot \mathbb{N}))$ and verify in PSPACE that the former is empty and the latter is not. Observe that now $\operatorname{gcd}(S) = p_1^{d_1} \cdots p_m^{d_i}$, meaning we can compute $\operatorname{gcd}(S)$ with polynomially many bits.

We now compute a bound $B \geq F(S)$. Let f_2 be the polynomial from Lemma 6.10. Since $S \cap (\mathbb{N} \setminus p_i^{d_i+1} \cdot \mathbb{N})$ is non-empty, Lemma 6.10 tells us that there is a number $n_i \in S \cap (\mathbb{N} \setminus p^{d_i+1} \cdot \mathbb{N})$ with $n_i \leq q^{f_2(n,k)}$. We can therefore guess a number $n_i \in \mathbb{N}$ with polynomially many digits and verify that $n_i \in S \cap (\mathbb{N} \setminus p_i^{d_i+1} \cdot \mathbb{N})$.

Since $p_i^{d_i+1}$ does not divide $n_i \in S$, we know that the set $T = \{r, n_1, \ldots, n_m\}$ satisfies $\gcd(T) = p_1^{d_1} \cdots p_m^{d_m} = \gcd(S)$. Therefore, the sets T^* and S^* are ultimately identical. Since trivially $T^* \subseteq S^*$, we may conclude $F(S) \le F(T)$. Moreover, according to Lemma 5.7, we have $F(T) \le (\operatorname{pmax}\{r, n_1, \ldots, n_m\})^2$ and we set $B := (\operatorname{pmax}\{r, n_1, \ldots, n_m\})^2$. Since $r \le q^{2f_1(n,k)}$ and $n_i \le q^{f_2(n,k)}$, we know that B is at most $q^{4f_1(n,k)+2f_2(n,k)}$ and can clearly be computed from r, n_1, \ldots, n_m . This completes the case $S \subseteq \mathbb{N}$.

In the case $S \subseteq -\mathbb{N}$, we can proceed analogously. If S contains a positive number and a negative number, we compute $\gcd(S)$ as above (replacing \mathbb{N} with \mathbb{Z}) and can set B=0 because F(S)=0.

7 Recognizability

In this section, we prove Theorem 3.4. We first present a characterization of recognizability that is easily checkable for PE-regular subsets. It is well-known that a subset S of \mathbb{Z} is recognizable if and only if there is a $k \in \mathbb{Z} \setminus \{0\}$ such that for every $s \in \mathbb{Z}$, we have $s \in S$ if and only if $s + k \in S$. Our characterization is an analog for Baumslag-Solitar groups.

A subset $S \subseteq \mathbb{Z}[\frac{1}{q}] \rtimes \mathbb{Z}$ is called k-periodic if for every $s \in \mathbb{Z}[\frac{1}{q}] \rtimes \mathbb{Z}$, we have (i) $s \in S$ if and only if $s(0,k) \in S$ and (ii) for every $\ell \in \mathbb{Z}$, we have $s \in S$ if and only if $s(q^{\ell} - q^{\ell + k}, 0) \in S$. In other words, membership in S is insensitive to (i) moving the cursor k positions and (ii) replacing a power of q by another power of q whose exponent differs by k. The set S is periodic if it is k-periodic for some $k \geq 1$. We show the following:

22

▶ Proposition 7.1. A subset $S \subseteq \mathbb{Z}[\frac{1}{q}] \times \mathbb{Z}$ is recognizable if and only if S is periodic.

Proof. Recall that S is k-periodic if

$$s \in S \iff s(0,k) \in S \quad \text{and} \quad s \in S \iff s(q^{\ell} - q^{\ell+k}, 0) \in S \text{ for every } \ell \in \mathbb{Z}.$$
 (5)

Suppose S is recognizable with a morphism $\varphi \colon \mathbb{Z}[\frac{1}{q}] \rtimes \mathbb{Z} \to K$ for some finite group K. Then there must be some $k \in \mathbb{Z} \setminus \{0\}$ with $\varphi((0,k)) = 1$: Otherwise, the map $\mathbb{Z} \to K$, $m \mapsto \varphi((0,m))$ would be injective, which is impossible for finite K. Now $\varphi((0,k)) = 1$ implies that $s \in S$ if and only if $s(0,k) \in S$ and thus the left equivalence in Equation (5). Moreover, since

$$(q^{\ell} - q^{\ell+k}, 0) = (q^{\ell}, 0)(0, k)(-q^{\ell}, 0)(0, -k),$$

we have $\varphi((q^{\ell}-q^{\ell+k},0))=1$ and hence S satisfies the right equivalence in Equation (5). Thus S is k-periodic.

Suppose S is k-periodic for $k \geq 1$ and consider the subgroup H of $G = \mathbb{Z}[\frac{1}{q}] \times \mathbb{Z}$ generated by (0,k) and by $(q^{\ell} - q^{\ell+k}, 0)$ for all $\ell \in \mathbb{Z}$. We claim that H is normal and the quotient G/H is finite. For normality, we have to check that for every generator h of H and every generator h of H on the following cases:

- Let h = (0, k) and g = (1, 0). Then $ghg^{-1} = (1, 0)(0, k)(-1, 0) = (q q^k, 0)$.
- Let h = (0, k) and g = (0, 1). Then $ghg^{-1} = (0, 1)(0, k)(0, -1) = (0, k)$.
- Let $h = (q^{\ell} q^{\ell+k}, 0)$ and g = (1,0). Then $ghg^{-1} = (1,0)(q^{\ell} q^{\ell+k}, 0)(-1,0) = (q^{\ell} q^{\ell+k}, 0)$.
- Let $h = (q^{\ell} q^{\ell+k}, 0)$ and g = (0,1). Then $ghg^{-1} = (0,1)(q^{\ell} q^{\ell+k}, 0)(0,-1) = (q^{\ell+1} q^{\ell+1+k}, 0)$.

In each case, ghg^{-1} clearly belongs to H, hence H is normal.

We may therefore consider the quotient group G/H and the projection $\pi\colon G\to G/H$. Note that since S is k-periodic, we know that for $s\in S$ if and only if $sh\in S$ for any $s\in S$ and $h\in H$. Therefore, if $\pi(s)=\pi(s')$, then $s\in S$ if and only $s'\in S$. Thus, S is recognized by the morphism π and it suffices to show that G/H is finite.

We prove this by showing that for any $(\frac{p}{q^\ell}, m) \in G$, we can multiply elements from H to obtain an element (r, n) with $r \in \{0, \pm 1, \pm 2, \dots, \pm q^k - 1\}$ and $n \in \{0, 1, \dots, k - 1\}$. Since there are only finitely many elements of the latter shape, this clearly implies finiteness of G/H. We do this in three steps. We first transform the left component into a natural number. Then we turn the left component into a number in $\{0, \pm 1, \pm 2, \dots, \pm q^k - 1\}$. Finally. we bring the right component to a number in $\{0, \dots, k - 1\}$.

For the first step, consider the element $g=(\frac{p}{q^\ell},m)\in G$. By multiplying $(-q^{-m-\ell}+q^{-m-\ell+k},0)^p\in H$ to g, we obtain $(\frac{p}{q^{\ell-k}},m)$. If we repeat this, we end up with an element (p,m) with $p\in\mathbb{Z}$ and $m\in\mathbb{Z}$.

For the second step, consider $(p,m) \in G$ with $p,m \in \mathbb{Z}$. If $p \geq q^k$, we multiply with $(1-q^k,0)$ and obtain $(p+1-q^k,0)$, where $p+1-q^k < p$ (because $k \geq 1$). By repeating this, we end up at an element (p,m) with $0 \leq p < q^k$. In the case $p < -q^k$, we just multiply $(-q+q^k,0) = (q-q^k,0)^{-1}$ instead of $(q-q^k,0)$. Thus, in general, we obtain an element (p,m) with $p \in \{0,\pm 1,\pm 2,\ldots,\pm q^k-1\}$.

For the third step, we merely reduce the right component modulo k: By multiplying (0, k) or (0, -k), we can clearly obtain an element (p, m) where $m \in \{0, 1, ..., k-1\}$ and where still $p \in \{0, \pm 1, \pm 2, ..., \pm q^k - 1\}$. Thus G/H is is finite and recognizability of S follows.

To decide whether a PE-regular $R \subseteq \mathrm{BS}(1,q)$ is recognizable, we show effective regularity of the set $N \subseteq \{\mathtt{a}\}^*$ of all words \mathtt{a}^k such that R is not k-periodic. Then, we just have to check whether N contains all words \mathtt{a}^k with $k \ge 1$, which is clearly decidable. Since R is PE-regular, the set $D = R(G \setminus R)^{-1} \cup (G \setminus R)R^{-1}$ is effectively PE-regular (Proposition 4.1). Then R is not k-periodic if and only if $(0,k) \in D$ or $(q^\ell - q^{\ell+k},0) \in D$ for some $\ell \in \mathbb{Z}$. The element (0,k) has the pointed expansion $0 \triangleleft 0^k - 10_\bullet$. The pointed expansions of $(q^\ell - q^{\ell+k},0)$ for $\ell \in \mathbb{Z}$ are exactly those words obtained from words $-0^r(q-1)^{k-1}0^s$ for $r,s \in \mathbb{N}$ by decorating one of the digits with \triangleleft and with \bullet , and removing leading or trailing 0's. Therefore, it is easy to see that $T_1 = \{(0 \triangleleft 0^{k-1}0_\bullet, \mathbf{a}^k) \mid k \ge 1\}$ and $T_2 = \{(\mathsf{pe}((q^\ell - q^{\ell+k},0)), \mathbf{a}^k) \mid \ell \in \mathbb{Z}, \ k \ge 1\}$ are rational transductions. This implies that $N = T_1(\mathsf{pe}(D)) \cup T_2(\mathsf{pe}(D)) \subseteq \mathbf{a}^*$ is effectively regular. Then clearly, R is not k-periodic if and only if $\mathbf{a}^k \in N$.

References

- Nathalie Aubrun and Jarkko Kari. Tiling problems on Baumslag-Solitar groups. In Proceedings of Machines, Computations and Universality 2013 (MCU 2013), pages 35–46, 2013. doi: 10.4204/EPTCS.128.12.
- 2 Laurent Bartholdi and Pedro V. Silva. Rational subsets of groups. CoRR, abs/1012.1532, 2010. Chapter 23 of the handbook AutoMathA (to appear). arXiv:1012.1532.
- 3 Gilbert Baumslag and Donald Solitar. Some two-generator one-relator non-Hopfian groups. *Bulletin of the American Mathematical Society*, 68(3):199–201, 1962. doi:10.1090/S0002-9904-1962-10745-9.
- Galina Aleksandrovna Bazhenova. Rational sets in finitely generated nilpotent groups. Algebra and Logic, 39(4):215–223, 2000. doi:10.1007/BF02681647.
- Jason P. Bell, Kathryn Hare, and Jeffrey Shallit. When is an automatic set an additive basis? Proceedings of the American Mathematical Society, Series B, 5(6):50-63, 2018. doi: 10.1090/bproc/37.
- 6 Michèle Benois. Parties rationnelles du groupe libre. CR Acad. Sci. Paris, 269:1188–1190, 1969.
- 7 Jean Berstel. Transductions and Context-Free Languages. Teubner, 1979.
- 8 Valérie Berthé and Michel Rigo, editors. Combinatorics, automata, and number theory, volume 135 of Encyclopedia of Mathematics and its Applications. Cambridge University Press, 2010.
- 9 Dmitry Chistikov, Wojciech Czerwiński, Piotr Hofman, Michal Pilipczuk, and Michael Wehar. Shortest paths in one-counter systems. Logical Methods in Computer Science, 15(1), 2019. doi:10.23638/LMCS-15(1:19)2019.
- Laura Ciobanu and Murray Elder. Solutions sets to systems of equations in hyperbolic groups are EDT0L in PSPACE. In Proceedings of the 46th International Colloquium on Automata, Languages, and Programming (ICALP 2019), pages 110:1-110:15, 2019. doi: 10.4230/LIPIcs.ICALP.2019.110.
- Jordi Delgado Rodríguez. Extensions of free groups: algebraic, geometric, and algorithmic aspects. PhD thesis, Universitat Politècnica de Catalunya. Facultat de Matemàtiques i Estadística, 2017.
- Volker Diekert, Claudio Gutierrez, and Christian Hagenah. The existential theory of equations with rational constraints in free groups is PSPACE -complete. *Information and Computation*, 202(2):105 140, 2005. doi:10.1016/j.ic.2005.04.002.
- Volker Diekert, Olga Kharlampovich, Markus Lohrey, and Alexei G. Myasnikov. Algorithmic Problems in Group Theory (Dagstuhl Seminar 19131). *Dagstuhl Reports*, 9(3):83–110, 2019. doi:10.4230/DagRep.9.3.83.
- Volker Diekert and Jürn Laun. On computing geodesics in Baumslag-Solitar groups. *International Journal on Algebra and Computation*, 21(1-2):119–145, 2011. doi:10.1142/S0218196711006108.

- Volker Diekert, Alexei G. Myasnikov, and Armin Weiß. Conjugacy in Baumslag's group, generic case complexity, and division in power circuits. In *Proceedings of 11th Latin American Symposium on Theoretical Informatics (LATIN 2014)*, pages 1–12, 2014. doi:10.1007/978-3-642-54423-1_1.
- Volker Diekert, Igor Potapov, and Pavel Semukhin. Decidability of membership problems for flat rational subsets of $GL(2,\mathbb{Q})$ and singular matrices, 2019. arXiv:1910.02302.
- 17 F. A. Dudkin and A. V. Treyer. Knapsack problem for Baumslag-Solitar groups. Siberian Journal of Pure and Applied Mathematics, 18:43-55, 2018. doi:10.33048/pam.2018.18.404.
- Murray Elder. A linear-time algorithm to compute geodesics in solvable Baumslag–Solitar groups. *Illinois Journal of Mathematics*, 54(1):109–128, 2010. doi:10.1215/ijm/1299679740.
- Murray Elder, Gillian Elston, and Gretchen Ostheimer. On groups that have normal forms computable in logspace. *Journal of Algebra*, 381:260 281, 2013. doi:10.1016/j.jalgebra. 2013.01.036.
- 20 Kousha Etessami, Dominik Wojtczak, and Mihalis Yannakakis. Quasi-birth-death processes, tree-like QBDs, probabilistic 1-counter automata, and pushdown systems. *Perform. Eval.*, 67(9):837–857, 2010.
- 21 Seymour Ginsburg and Edwin H. Spanier. Bounded regular sets. *Proceedings of the American Mathematical Society*, 17(5):1043–1049, 1966. doi:10.2307/2036087.
- Oscar H. Ibarra. Reversal-bounded multicounter machines and their decision problems. *Journal of the ACM*, 25(1):116–133, 1978. doi:10.1145/322047.322058.
- 23 Ilya Kapovich and Alexei Myasnikov. Stallings foldings and subgroups of free groups. *Journal of Algebra*, 248(2):608 668, 2002. doi:10.1006/jabr.2001.9033.
- 24 Olga Kharlampovich, Laura López, and Alexei Miasnikov. Diophantine problem in some metabelian groups, 2019. arXiv:1903.10068.
- Dexter Kozen. Lower bounds for natural proof systems. In *Proceedings of the 18th Annual Symposium on Foundations of Computer Science (FOCS 1977)*, pages 254–266, 1977. doi: 10.1109/SFCS.1977.16.
- 26 Michel Latteux. Langages à un compteur. J. Comput. Syst. Sci., 26(1):14-33, 1983.
- 27 Markus Lohrey. The rational subset membership problem for groups: a survey. In C. M. Campbell, M. R. Quick, E. F. Robertson, and C. M. Roney-Dougal, editors, *Groups St Andrews 2013*, volume 422 of *Lond. Math. S.*, pages 368–389, Cambridge, United Kingdom, 2016. Cambridge University Press. doi:10.1017/CB09781316227343.024.
- Markus Lohrey and Géraud Sénizergues. Rational subsets in HNN-extensions and amalgamated products. *International Journal on Algebra and Computation*, 18(1):111–163, 2008. doi: 10.1142/S021819670800438X.
- 29 Markus Lohrey and Georg Zetzsche. Knapsack in metabelian Baumslag-Solitar groups, 2020. arXiv:2002.03837.
- 30 Alexei Myasnikov, Andrey Nikolaev, and Alexander Ushakov. Knapsack problems in groups. Mathematics of Computation, 84:987–1016, 2015. doi:10.1090/S0025-5718-2014-02880-9.
- 31 David Robinson. Parallel Algorithms for Group Word Problems. PhD thesis, Department of Mathematics, University of Califoria, San Diego, 1993.
- 32 N. S. Romanovskii. Some algorithmic problems for solvable groups. Algebra and Logic, 13:13-16, 1974. doi:10.1007/BF01462922.
- N. S. Romanovskiĭ. The occurrence problem for extensions of abelian groups by nilpotent groups. Siberian Mathematical Journal, 21:273–276, 1980. doi:10.1007/BF00968275.
- 34 Géraud Sénizergues. On the rational subsets of the free group. Acta Informatica, 33(3):281–296, 1996. doi:10.1007/s002360050045.
- John C Shepherdson. The reduction of two-way automata to one-way automata. *IBM Journal of Research and Development*, 3(2):198–200, 1959. doi:10.1147/rd.32.0198.
- Pedro V. Silva. Free group languages: Rational versus recognizable. *RAIRO—Theoretical Informatics and Applications*, 38(1):49–67, 2004. doi:10.1051/ita:2004003.

- 37 Pedro V. Silva. An automata-theoretic approach to the study of fixed points of endomorphisms. In Ventura E. González-Meneses J., Lustig M., editor, *Algorithmic and Geometric Topics Around Free Groups and Automorphisms*, Advanced Courses in Mathematics—CRM Barcelona, pages 1–42. Birkhäuser, 2017. doi:10.1007/978-3-319-60940-9_1.
- 38 Armin Weiß. On the Complexity of Conjugacy in Amalgamated Products and HNN Extensions. PhD thesis, Institut für Formale Methoden der Informatik, Universität Stuttgart, 2015.
- 39 Herbert S. Wilf. A circle-of-lights algorithm for the "money-changing problem". *The American Mathematical Monthly*, 85(7):562–565, 1978. doi:10.2307/2320864.