Automatic Sets of Rational Numbers

Eric Rowland
Université de Liège
Département de Mathématiques
Grande Traverse 12 (B37)
4000 Liège
Belgium
erowland@ulg.ac.be

Jeffrey Shallit*
School of Computer Science
University of Waterloo
Waterloo, ON N2L 3G1
Canada
shallit@cs.uwaterloo.ca

February 24, 2014

Abstract

The notion of a k-automatic set of integers is well-studied. We develop a new notion — the k-automatic set of rational numbers — and prove basic properties of these sets, including closure properties and decidability.

1 Introduction

Let k be an integer ≥ 2, and let $\mathbb{N} = \{0, 1, 2, \ldots\}$ denote the set of non-negative integers. Let $\Sigma_k = \{0, 1, \ldots, k-1\}$ be an alphabet of k letters. Given a word $w = a_1 a_2 \cdots a_t \in \Sigma_k^*$, we let $[w]_k$ denote the integer that it represents in base k; namely,

$$[w]_k = \sum_{1 \leq i \leq t} a_i k^{t-i}, \tag{1}$$

*Corresponding author.

where (as usual) an empty sum is equal to 0. For example, $[101011]_2 = 43$.

Note that in this framework, every element of $\mathbb{N}$ has infinitely many distinct representations as words, each having a certain number of *leading zeroes*. Among all such representations, the one with no leading zeroes is called the *canonical representation*; it is an element of $C_k := \{\epsilon\} \cup (\Sigma_k - \{0\})\Sigma_k^*$. For an integer $n \geq 0$, we let $(n)_k$ denote its canonical representation. Note that ϵ is the canonical representation of 0.

Given a language $L \subseteq \Sigma_k^*$, we can define the set of integers it represents, as follows:

$$[L]_k = \{[w]_k \,:\, w \in L\}. \tag{2}$$

We now recall a well-studied concept, that of k-automatic set (see, e.g., [8, 9, 2]):

Definition 1. We say that a set $S \subseteq \mathbb{N}$ is *k-automatic* if there exists a regular language $L \subseteq \Sigma_k^*$ such that $S = [L]_k$.

Many properties of these sets are known. For example, it is possible to state an equivalent definition involving only canonical representations:

Definition 2. A set $S \subseteq \mathbb{N}$ is *k-automatic* if the language

$$(S)_k := \{(n)_k \,:\, n \in S\}$$

is regular.

To see the equivalence of Definitions 1 and 2, note that if L is a regular language, then so is the language L' obtained by removing all leading zeroes from each word in L.

A slightly more general concept is that of k-automatic *sequence*. Let Δ be a finite alphabet. Then a sequence (or infinite word) $(a_n)_{n\geq 0}$ over Δ is said to be k-automatic if, for every $c \in \Delta$, the set of *fibers* $F_c = \{n \in \mathbb{N} \,:\, a_n = c\}$ is a k-automatic set of natural numbers. Again, this class of sequences has been widely studied [8, 9, 2]. The following result is well-known [10]:

Theorem 3. *The sequence $(a_n)_{n\geq 0}$ is k-automatic if and only if its k-kernel, the set of its subsequences $K = \{(a_{k^e n+f})_{n\geq 0} \,:\, e \geq 0,\ 0 \leq f < k^e\}$, is finite.*

In previous papers [23, 22], the second author extended the notion of k-automatic sets over $\mathbb{N}$ to subsets of $\mathbb{Q}^{\geq 0}$, the non-negative rational numbers. The motivation was to study the "critical exponent" of automatic sequences. In this paper, we will obtain some basic results about this new class. Our principal results are Theorem 17 (characterizing those k-automatic sets of rationals consisting entirely of integers), Theorem 22 and Corollary 23 (showing that the class of k-automatic sets of rationals is not closed under intersection or complement), Theorem 29 (showing that it is decidable if a k-automatic set of rationals is infinite), and Theorem 33 (showing that it is decidable if a k-automatic set of rationals equals $\mathbb{N}$).

The class of sets we study has some similarity to another class studied by Even [11] and Hartmanis and Stearns [13]; their class corresponds to the topological closure of a small subclass of our k-automatic sets, in which the possible denominators are restricted to powers of k.

Yet another model of automata accepting real numbers was studied in [1, 4, 5, 6]. In this model real numbers are represented by their (possibly infinite) base-k expansions, and the model of automaton used is a nondeterministic Büchi automaton. However, even when restricted to rational numbers, this model does not define the same class of sets, as we will show below in Corollary 25.

A preliminary version of this paper appeared in [17].

2 Representing rational numbers

A natural representation for the non-negative rational number p/q is the pair (p, q) with $q \neq 0$. Of course, this representation has the drawback that every element of $\mathbb{Q}^{\geq 0}$ has infinitely many representations, each of the form $(jp/d, jq/d)$ for some $j \geq 1$, where $d = \gcd(p, q)$.

We might try to ensure uniqueness of representations by considering only "reduced" representations (those in "lowest terms"), which amounts to representing p/q by the pair $(p/d, q/d)$ where $d = \gcd(p, q)$. In other words, the only valid pairs are (p, q) with $\gcd(p, q) = 1$. However, the condition $\gcd(p, q) = 1$ cannot be checked, in general, by finite or even pushdown automata — see Remark 26 below — and it is not currently known if it is decidable whether a given regular language consists entirely of reduced representations (see Section 7). Furthermore, insisting on only reduced representations means that some "reasonable" sets of rationals, such as $\{(k^m - 1)/(k^n - 1) : m, n \geq 1\}$ (see Corollary 24), have no representation as a regular language. For these reasons, **we allow the rational number p/q to be represented by *any* pair of non-negative integers (p', q') with $p/q = p'/q'$.**

Next, we need to see how to represent a pair of integers as a word over a finite alphabet. Here, we follow the ideas of Salon [18, 19, 20]. Consider the alphabet Σ_k^2. A finite word w over Σ_k^2 can be considered as a sequence of pairs $w = [a_1, b_1][a_2, b_2] \cdots [a_n, b_n]$ where $a_i, b_i \in \Sigma_k$ for $1 \leq i \leq n$. We can now define the projection maps π_1, π_2 from $(\Sigma_k^2)^*$ to Σ_k^*, as follows:

$$\pi_1(w) = a_1 a_2 \cdots a_n; \qquad \pi_2(w) = b_1 b_2 \cdots b_n.$$

Then we define $[w]_k = ([\pi_1(w)]_k, [\pi_2(w)]_k)$. Thus, for example, if

$$w = [0, 0][1, 0][0, 1][1, 0][0, 0][1, 1][1, 0],$$

then $[w]_2 = (43, 18)$. We also define $\times$, which allows us to join two words $w, x \in \Sigma_k^*$ of the same length to create a single word $w \times x \in (\Sigma_k^2)^*$ whose π_1 projection is w and π_2 projection is x.

In this framework, every pair of integers (p, q) again has infinitely many distinct representations, arising from padding on the left by leading pairs of zeroes, that is, by $[0, 0]$. Among all such representations, the *canonical representation* is the one having no leading pairs of zeroes. We write it as $(p, q)_k$. For example, $(43, 18)_2 = [1, 0][0, 1][1, 0][0, 0][1, 1][1, 0]$.

We now state the fundamental definitions of this paper:

Definition 4. Given a word $w \in (\Sigma_k^2)^*$ with $[\pi_2(w)]_k \neq 0$, we define

$$\mathrm{quo}_k(w) := \frac{[\pi_1(w)]_k}{[\pi_2(w)]_k}.$$

If $[\pi_2(w)]_k = 0$, then $\mathrm{quo}_k(w)$ is not defined. Further, if $[\pi_2(w)]_k \neq 0$ for all $w \in L$, then $\mathrm{quo}_k(L) := \{\mathrm{quo}_k(w) : w \in L\}$. A set of rational numbers $S \subseteq \mathbb{Q}^{\geq 0}$ is *k-automatic* if there exists a regular language $L \subseteq (\Sigma_k^2)^*$ such that $S = \mathrm{quo}_k(L)$.

We reiterate that for a rational number α to be in $\mathrm{quo}_k(L)$, only a single, possibly non-reduced, representation of α need be in L. Furthermore, L may contain multiple representations for α in two different ways: L could contain non-canonical representations that begin with leading zeroes, and L could contain "unreduced" representations (p, q) where $\gcd(p, q) > 1$.

We could have adopted a different definition, by insisting that every rational in $\mathrm{quo}_k(L)$ must have every possible representation in L. However, this would have the unpleasant consequence that some very simple subsets of $\mathbb{Q}^{\geq 0}$ would not have a regular representation. For example, the language of all representations of $\mathbb{N}$ is given by $L_d = \{(p, q)_k : q \mid p\}$, which is not even context-free; see Remark 26.

Another issue is the following: given a set $S \subseteq \mathbb{Q}^{\geq 0}$, if S contains a non-integer, then by calling S k-automatic, it is clear that we intend this to mean that S is k-automatic in the sense of Definition 4. But what if $S \subseteq \mathbb{N}$? Then calling it "automatic" might mean *either* automatic in the usual sense, as in Definition 1, *or* in the extended sense introduced in this section, treating S as a subset of $\mathbb{Q}^{\geq 0}$. In Theorem 17 we will see that these two interpretations actually *coincide* for subsets of $\mathbb{N}$, but in order to prove this, we need some notation to distinguish between the two types of representations. So, from now on, by $(\mathbb{N}, k)$-automatic we mean the interpretation in Definition 1, and by $(\mathbb{Q}^{\geq 0}, k)$-automatic we mean the interpretation in Definition 4.

Yet another issue is the order of the representations. So far we have only considered representations where the leftmost digit is the most significant digit. However, sometimes it is simpler to deal with *reversed representations* where the leftmost digit is the least significant digit. In other words, sometimes it is easier to deal with the reversed word w^R and reversed language L^R instead of w and L, respectively. Since the regular languages are (effectively) closed under reversal, for most of our results it does not matter which representation we choose, and we omit extended discussion of this point.

We use the following notation for intervals: $I[\alpha, \beta]$ denotes the closed interval

$$\{x : \alpha \leq x \leq \beta\},$$

and similarly for open- and half-open intervals.

3 Examples

To build intuition, we give some examples of k-automatic sets of rationals.

Example 5. Consider the regular language

$$L_0 = \{w \in (\Sigma_k^2)^* \; : \; \pi_1(w) \in C_k \; \cup \; \{0\} \text{ and } [\pi_2(w)]_k = 1\}.$$

Then $\text{quo}_k(L_0) = \mathbb{N}$, as L_0 contains words with arbitrary numerator, but denominator equal to 1.

Example 6. Let $k = 2$, and consider the regular language L_1 defined by the regular expression $A^*\{[0,1],[1,1]\}A^*$, where $A = \{[0,0],[0,1],[1,0],[1,1]\}$. This regular expression specifies all pairs of integers where the second component has at least one nonzero digit — the point being to avoid division by 0. Then $\text{quo}_k(L_1) = \mathbb{Q}^{\geq 0}$, the set of all non-negative rational numbers. In fact all possible representations of all rational numbers are included in L_1.

Example 7. Consider the regular language

$$L_2 = \{w \in (\Sigma_k^2)^* \; : \; \pi_2(w) \in 0^*1^+0^*\}.$$

Then we claim that $\text{quo}_k(L_2) = \mathbb{Q}^{\geq 0}$. To see this, consider an arbitrary non-negative rational number p/q with $q \geq 1$. Let i be the least non-negative integer such that $\gcd(k^i, q) = \gcd(k^{i+1}, q)$. Let $d = \gcd(k^i, q)$ and write $q = dq'$; note that $\gcd(k, q') = 1$ and hence $\gcd(k, (k-1)q') = 1$. By the Fermat-Euler theorem, there is an integer $j \geq 1$ such that $k^j \equiv 1 \pmod{(k-1)q'}$. (For example, we can take $j = \varphi((k-1)q')$, where φ is Euler's totient function.) Define $d' = k^i/d$ and $t = \frac{k^j-1}{(k-1)q'}$; then

$$\frac{p}{q} = \frac{d'tp}{d'tq} = \frac{d'tp}{k^i \frac{k^j-1}{k-1}},$$

which expresses p/q as a number with denominator of the required form, with base-k representation $1^j\, 0^i$. In Theorems 35 and 36 below we will see that L_2 achieves the minimum subword complexity of $\pi_2(L)$ over all regular languages L representing $\mathbb{Q}^{\geq 0}$.

Example 8. For a word x and letter a let $|x|_a$ denote the number of occurrences of a in x. Consider the regular language

$$L_3 = \{w \in (\Sigma_2^2) \; : \; |\pi_1(w)|_1 \equiv 0 \pmod 2 \text{ and } |\pi_2(w)|_1 \equiv 1 \pmod 2\}.$$

Then it follows from a result of Schmid [21] that $\text{quo}_2(L_3) = \mathbb{Q}^{\geq 0} - \{2^n \; : \; n \in \mathbb{Z}\}$.

Example 9. Let $k = 3$, and consider the regular language L_4 defined by the regular expression $[0,1]\{[0,0],[2,0]\}^*$. Then $\text{quo}_k(L_4)$ is the 3-*adic Cantor set*, the set of all rational numbers in the "middle-thirds" Cantor set with denominators a power of 3 [7].

Example 10. Let $k = 2$, and consider the regular language L_5 defined by the regular expression $[0,1]\{[0,0],[0,1]\}^*\{[1,0],[1,1]\}$. Then the numerator encodes the integer 1, while the denominator encodes all integers that start with 1. Hence $\text{quo}_k(L_5) = \{\frac{1}{n} \; : \; n \geq 1\}$.

Example 11. Let $k = 4$, and consider the set $S = \{0, 1, 3, 4, 5, 11, 12, 13, \ldots\}$ of all non-negative integers that can be represented using only the digits $0, 1, -1$ in base 4. Consider the set $L_6 = \{(p, q)_4 \ : \ p, q \in S\}$. It is not hard to see that L_6 is $(\mathbb{Q}^{\geq 0}, 4)$-automatic. The main result in [15] can be phrased as follows: $\mathrm{quo}_4(L_6)$ contains every odd integer. In fact, an integer t is in $\mathrm{quo}_4(L_6)$ if and only if the exponent of the largest power of 2 dividing t is even.

Example 12. Let K, L be arbitrary regular languages over the alphabet Σ_k^2. Note that $\mathrm{quo}_k(K \cup L) = \mathrm{quo}_k(K) \cup \mathrm{quo}_k(L)$ but the analogous identity involving intersection need not hold. As an example, consider $K = \{[2, 1]\}$ and $L = \{[4, 2]\}$. Then $\mathrm{quo}_{10}(K \cap L) = \emptyset \neq \{2\} = \mathrm{quo}_{10}(K) \cap \mathrm{quo}_{10}(L)$.

4 Basic results

In this section we obtain some basic results about automatic sets of rationals.

Theorem 13. *Let r, s be integers ≥ 1. Then S is a k^r-automatic set of rational numbers if and only if S is k^s-automatic.*

Proof. Follows easily from the same result for automatic sequences [8, 9]. □

Next we state a useful result from [23, 22]:

Lemma 14. *Let β be a non-negative real number and define the languages*

$$L_{\leq \beta} = \{x \in (\Sigma_k^2)^* \ : \ \mathrm{quo}_k(x) \leq \beta\},$$

and analogously for the relations $<, =, \geq, >, \neq$.

(a) If β is a rational number, then the language $L_{\leq \beta}$ (resp., $L_{<\beta}$, $L_{=\beta}$, $L_{\geq \beta}$, $L_{>\beta}$, $L_{\neq \beta}$) is regular.

(b) If $L_{\leq \beta}$ (resp., $L_{<\beta}$, $L_{\geq \beta}$, $L_{>\beta}$) is regular, then β is a rational number.

Suppose S is a set of real numbers, and α is a real number. We introduce the following notation:

$$\begin{aligned} S + \alpha &:= \{x + \alpha \ : \ x \in S\} \\ S \dot{-} \alpha &:= \{\max(x - \alpha, 0) \ : \ x \in S\} \\ \alpha \dot{-} S &:= \{\max(\alpha - x, 0) \ : \ x \in S\} \\ \alpha S &:= \{\alpha x \ : \ x \in S\}. \end{aligned}$$

Theorem 15. *The class of k-automatic sets of rational numbers is closed under the following operations:*

(i) *union;*

(ii) $S \to S + \alpha$ *for* $\alpha \in \mathbb{Q}^{\geq 0}$*;*

(iii) $S \to S \dot{-} \alpha$ *for* $\alpha \in \mathbb{Q}^{\geq 0}$*;*

(iv) $S \to \alpha \dot{-} S$ *for* $\alpha \in \mathbb{Q}^{\geq 0}$*;*

(v) $S \to \alpha S$ *for* $\alpha \in \mathbb{Q}^{\geq 0}$*;*

(vi) $S \to \{1/x \, : \, x \in S \setminus \{0\}\}$.

Proof. We prove only item (ii), with the others being similar. We will use the reversed representation, with the least significant digit appearing first. Write $\alpha = p/q$. Let M be a DFA with $\text{quo}_k(L(M)) = S$. To accept $S + \alpha$, on input a base-k representation of $x = p'/q'$, we transduce the numerator to $p'q - pq'$ and the denominator to qq' (hence effectively computing a representation for $x - \alpha$), and simultaneously simulate M on this input, digit-by-digit, accepting if M accepts. $\square$

The next theorem shows that if p/q has a representation in a regular language, then it has a representation where the numerator and denominator are not too large. Here we are using the least-significant-digit first representation.

Theorem 16. *Let* $L \subseteq (\Sigma_k^2)^*$ *be a regular language, accepted by an NFA* M *with* n *states. If* $p/q \in \text{quo}_k(L)$*, there exists* $w \in L$ *with* $\pi_1(w) = p'$*,* $\pi_2(w) = q'$*,* $p'/q' = p/q$*, and* $p', q' < k^{pqn}$*.*

Proof. On input w representing the pair (p', q') we compute qp' and pq' simultaneously, digit-by-digit, and test if they are equal. To carry out multiplication by q, we need to keep track of carries, which could be as large as $q - 1$, and similarly for p. An NFA M' to do this can be built using triples (i, j, q_l), where i is a carry in the multiplication by p, j is a carry in the multiplication by q, and q_l is the state in M arising from processing a prefix of w. Thus M' has pqn states. If M' accepts any word, then it accepts a word of length at most $pqn - 1$. From this the inequality follows. $\square$

We now state one of our main results.

Theorem 17. *Let* $S \subseteq \mathbb{N}$*. Then* S *is* $(\mathbb{N}, k)$*-automatic if and only if it is* $(\mathbb{Q}^{\geq 0}, k)$*-automatic.*

The proof requires a number of preliminary results. First, we introduce some terminology and notation. We say a set $S \subseteq \mathbb{N}$ is *ultimately periodic* if there exist integers $n_0 \geq 0, p \geq 1$ such that $n \in S \iff n + p \in S$, provided $n \geq n_0$. In particular, every finite set is ultimately periodic.

We let $\mathcal{P} = \{2, 3, 5, \ldots\}$ denote the set of prime numbers. Given a positive integer n, we let $\text{pd}(n)$ denote the set of its prime divisors. For example, $\text{pd}(60) = \{2, 3, 5\}$. Given a subset $D \subseteq \mathcal{P}$, we let $\pi(D) = \{n \geq 1 \, : \, \text{pd}(n) \subseteq D\}$, the set of all integers that can be factored completely using only the primes in D. Finally, recall that $\nu_k(n)$ denotes the exponent of the largest power of k dividing n.

First, we prove two useful lemmas.

Lemma 18. *Let $S \subseteq \mathbb{N} - \{0\}$. Then the following are equivalent:*

(a) There exist an integer $n \geq 0$, and n integers $g_i \geq 1$, $1 \leq i \leq n$, and n ultimately periodic subsets $W_i \subseteq \mathbb{N}$, $1 \leq i \leq n$, such that

$$S = \bigcup_{1 \leq i \leq n} g_i \{k^j \ : \ j \in W_i\};$$

(b) There exist an integer $m \geq 0$, and m integers f_i with $k \not| f_i$, $1 \leq i \leq m$, and m ultimately periodic subsets $V_i \subseteq \mathbb{N}$, $1 \leq i \leq m$, such that

$$S = \bigcup_{1 \leq i \leq m} f_i \{k^j \ : \ j \in V_i\}, \tag{3}$$

and the union is disjoint.

(c) Define $F = \{s/k^{\nu_k(s)} \ : \ s \in S\}$. The set F is finite, and for all $f \in F$, the set $U_f = \{j \ : \ k^j f \in S\}$ is ultimately periodic.

Proof. (a) $\Longrightarrow$ (b): For each g_i define $x_i = \nu_k(g_i)$ and $f_i = g_i/k^{x_i}$. Note that $k \not| f_i$. Then

$$\begin{aligned} S &= \bigcup_{1 \leq i \leq n} g_i \{k^j \ : \ j \in W_i\} \\ &= \bigcup_{1 \leq i \leq n} f_i k^{x_i} \{k^j \ : \ j \in W_i\} \\ &= \bigcup_{1 \leq i \leq n} f_i \{k^{x_i+j} \ : \ j \in W_i\} \\ &= \bigcup_{1 \leq i \leq n} f_i \{k^j \ : \ j \in W_i'\}, \end{aligned}$$

where $W_i' = x_i + W_i$. Note that each W_i' is ultimately periodic. If any of the f_i coincide, we take the union of the corresponding W_i' and call it V_i. Since the union of a finite number of ultimately periodic sets is still ultimately periodic (see, e.g., [16]), we can choose a subset of the indices i so that each f_i appears once, expressing S as

$$\bigcup_{1 \leq i \leq m} f_i \{k^j \ : \ j \in V_i\}$$

for some $m \leq n$. The union is now disjoint, for if (say) $f_1 k^j = f_2 k^{j'}$ for $j \in V_1$ and $j' \in V_2$, then since $k \not| f_1, f_2$ we must have $j = j'$ and so $f_1 = f_2$, which is a contradiction.

(b) $\Longrightarrow$ (c):

Let $s \in S$. Then $s = f_i k^j$ for some j. Since $k \not| f_i$, we have $s/k^{\nu_k(s)} = f_i$. Thus $F = \{f_i \ : \ 1 \leq i \leq m\}$ and hence is finite. By the disjointness of the union (3) (and the fact that $k \not| f_i$) we have $k^j f_i \in S \iff j \in V_i$. So $U_{f_i} = V_i$ and hence U_{f_i} is ultimately periodic.

(c) $\Longrightarrow$ (a):

Let $g_1, g_2, \ldots, g_n$ be distinct elements such that $F = \{g_1, g_2, \ldots, g_n\}$. Take $W_i = U_{g_i}$. $\square$

If any of the conditions (a)–(c) above hold, we say that the set S is *k-finite*.

Lemma 19. *Let D be a finite set of prime numbers, and let $S \subseteq \pi(D)$. Let $s_1, s_2, \ldots$ be an infinite sequence of (not necessarily distinct) elements of S. Then there is an infinite increasing sequence of indices $i_1 < i_2 < \cdots$ such that $s_{i_1} \mid s_{i_2} \mid \cdots$.*

Proof. Case 1: The sequence (s_i) is bounded. In this case infinitely many of the s_i are the same, so we can take the indices to correspond to these s_i.

Case 2: The sequence (s_i) is unbounded. In this case we prove the result by induction on $|D|$. If $|D| = 1$, then we can choose a strictly increasing subsequence of the (s_i); since all are powers of some prime p, this subsequence has the desired property.

Now suppose the result is true for all sets D of cardinality $t-1$. We prove it for $|D| = t$. Since only t distinct primes figure in the factorization of the s_i, some prime, say p, must appear with unbounded exponent in the (s_i). So there is some subsequence of (s_i), say (t_i), with strictly increasing exponents of p. Now consider the infinite sequence (u_i) given by $u_i = t_i / p^{\nu_p(t_i)}$. Each u_i has a prime factorization in terms of the primes in $D - \{p\}$, so by induction there is an infinite increasing sequence of indices $i_1, i_2, \ldots$ such that $u_{i_1} \mid u_{i_2} \mid \cdots$. Then $p^{\nu_p(t_{i_1})} u_{i_1} \mid p^{\nu_p(t_{i_2})} u_{i_2} \mid \cdots$, which corresponds to an infinite increasing sequence of indices of the original sequence (s_i). $\square$

We now state an essential part of the proof, which is of independent interest.

Theorem 20. *Let $D \subseteq \mathcal{P}$ be a finite set of prime numbers, and let $S \subseteq \pi(D)$. Then S is k-automatic if and only if it is k-finite.*

Proof. $\Longleftarrow$: If S is k-finite, then by Lemma 18, we can write it as the disjoint finite union

$$S = \bigcup_{1 \le i \le m} f_i \{k^j \ : \ j \in V_i\},$$

where each V_i is an ultimately periodic set of integers. For each i, the set $(f_i)_k \{0^t \ : \ t \in V_i\}$ of base-k representations of $f_i\{k^j \ : \ j \in V_i\}$ is a regular language. It follows that S is k-automatic.

$\Longrightarrow$: Suppose S is k-automatic. Define $F = \{s/k^{\nu_k(s)} \ : \ s \in S\}$. Suppose F is infinite. Clearly no element of F is divisible by k. Therefore we can write S as the disjoint union $\bigcup_{i \ge 0} k^i H_i$, where $H_i := \{f \in F \ : \ k^i f \in S\}$. Then there are two possibilities: either (a) the sets H_i are finite for all $i \ge 0$, or (b) at least one H_i is infinite.

In case (a), define $u_i := \max H_i$ for all $i \ge 0$. Then the set $\{u_0, u_1, \ldots\}$ must be infinite, for otherwise F would be finite. Choose an infinite subsequence of the u_i consisting of distinct elements, and apply Lemma 19. Then there is an infinite increasing subsequence of indices $i_1 < i_2 < \cdots$ such that $u_{i_1} \mid u_{i_2} \mid \cdots$. So the sequence $(u_{i_j})_{j \ge 1}$ is strictly increasing.

Now consider the characteristic sequence of S, say $(f(n))_{n \ge 0}$, taking values 1 if $n \in S$ and 0 otherwise. Consider the subsequences (f_j) in the k-kernel of f defined by $f_j(n) = f(k^{i_j} n)$ for $n \ge 0$, $j \ge 1$. By our construction, the largest n in $\pi(D)$ such that $k \not| n$ and $f_j(n) = 1$ is

$n = u_{i_j}$. Since the u_{i_j} are strictly increasing, this shows the (infinitely many) sequences (f_j) are pairwise distinct. Hence, by Theorem 3, f is not k-automatic and neither is S.

In case (b), we have H_i is infinite for some i. As mentioned above, S can be written as the disjoint union $\bigcup_{i\geq 0} k^i H_i$. Let L be the language of canonical base-k expansions of elements of H_i (so that, in particular, no element of L starts with 0). The base-k representation of elements of $k^i H_i$ end in exactly i 0's, and no other member of S has this property. Since S is assumed to be k-automatic, it follows that L is regular. Note that no two elements of H_i have a quotient which is divisible by k, because if they did, the numerator would be divisible by k, which is ruled out by the condition.

Since L is infinite and regular, by the pumping lemma, there must be words u, v, w, with v nonempty, such that $uv^j w \in L$ for all $j \geq 0$. Note that for all integers $j \geq 0$ and $c \geq 0$ we have

$$[uv^{j+c}w]_k = [uv^j w]_k \cdot k^{c|v|} + ([v^c w]_k - [w]_k \cdot k^{c|v|}). \tag{4}$$

Let $D = \{p_1, p_2, \ldots, p_t\}$. Since $[uv^j w]_k \in S \subseteq \pi(D)$, it follows that there exists a double sequence $(f_{r,j})_{1\leq r\leq t; j\geq 1}$ of non-negative integers such that

$$[uv^j w]_k = p_1^{f_{1,j}} \cdots p_t^{f_{t,j}} \tag{5}$$

for all $j \geq 0$. From (5), we see that $k^{|uw|+j|v|} < p_1^{f_{1,j}} \cdots p_t^{f_{t,j}}$, and hence (assuming $p_1 < p_2 < \cdots < p_t$) we get

$$(|uw| + j|v|) \log k < \left(\sum_{1\leq r\leq t} f_{r,j} \right) \log p_t.$$

Therefore, there are constants $0 < c_1$ and J such that $c_1 j < \sum_{1\leq r\leq t} f_{r,j}$ for $j \geq J$.

For each $j \geq J$ now consider the indices r such that $f_{r,j} > c_1 j/t$; there must be at least one such index, for otherwise $f_{r,j} \leq c_1 j/t$ for each r and hence $\sum_{1\leq r\leq t} f_{r,j} \leq c_1 j$, a contradiction. Now consider $t+1$ consecutive j's; for each j there is an index r with $f_{r,j} > c_1 j/t$, and since there are only t possible indices, there must be an r and two integers l' and l, with $0 \leq l < l' \leq t$, such that $f_{r,j+l} > c_1(j+l)/t$ and $f_{r,j+l'} > c_1(j+l')/t$. This is true in any block of $t+1$ consecutive j's that are $\geq J$. Now there are infinitely many disjoint blocks of $t+1$ consecutive j's, and so there must be some pair $(r, l'-l)$ that occurs infinitely often. Put $\delta = l' - l$.

Now use (4) and take $c = \delta$. We get infinitely many j such that

$$p_1^{f_{1,j+\delta}} \cdots p_t^{f_{t,j+\delta}} = k^{\delta|v|} p_1^{f_{1,j}} \cdots p_t^{f_{t,j}} + E,$$

where $E = [v^\delta w]_k - [w]_k \cdot k^{\delta|v|}$ is a constant that is independent of j. Now focus attention on the exponent of p_r on both sides. On the left it is $f_{r,j+\delta}$, which we know to be at least $c_1(j+\delta)/t$. On the right the exponent of p_r dividing the first term is $f_{r,j} + \delta|v|e_r$ (where $k = p_1^{e_1} \cdots p_t^{e_t}$); this is at least $c_1 j/t$. So p_r^h divides E, where $h \geq c_1 j/t$. But this quantity goes to ∞, and E is a constant. So $E = 0$. But then

$$\frac{[uv^{j+\delta}w]_k}{[uv^j w]_k} = k^{\delta|v|}.$$

which is impossible, since, as we observed above, two elements of H_i cannot have a quotient which is a power of k. This contradiction shows that H_i cannot be infinite.

So now we know that F is finite. Fix some $f \in F$ and consider $T_f = \{k^i : k^i f \in S\}$. Since S is k-automatic, and the set of base-k expansions $(T_f)_k$ is essentially formed by stripping off the bits corresponding to $(f)_k$ from the front of each element of S of which $(f)_k$ is a prefix, and replacing it with "1", this is just a left quotient followed by concatenation, and hence $(T_f)_k$ is regular. Let M' be a DFA for $(T_f)_k$, and consider an input of 1 followed by l 0's for $l = 0, 1, \ldots$ in M'. Evidently we eventually get into a cycle, so this says that $U_f = \{i : k^i f \in S\}$ is ultimately periodic. Hence S is k-finite.

This completes the proof of Theorem 20. $\square$

We can now prove Theorem 17.

Proof. One direction is easy, since if S is $(\mathbb{N}, k)$-automatic, then there is an automaton accepting $(S)_k$. We can now easily modify this automaton to accept all words over $(\Sigma_k^2)^*$ whose π_2 projection represents the integer 1 and whose π_1 projection is an element of $(S)_k$. Hence S is $(\mathbb{Q}^{\geq 0}, k)$-automatic.

Now assume $S \subseteq \mathbb{N}$ is $(\mathbb{Q}^{\geq 0}, k)$-automatic. If S is finite, then the result is clear, so assume S is infinite. Let L be a regular language with $\mathrm{quo}_k(L) = S$. Without loss of generality we may assume every representation in L is canonical; there are no leading $[0, 0]$'s. Furthermore, by first intersecting with $L_{\neq 0}$ we may assume that L contains no representations of the integer 0. Finally, we can also assume, without loss of generality, that no representation contains *trailing* occurrences of $[0, 0]$, for removing all trailing $[0, 0]$'s from all words in L preserves regularity, and it does not change the set of numbers represented, as it has the effect of dividing the numerator and denominator by the same power of k. Since the words in L represent integers only, the denominator of every representation must divide the numerator, and hence if the denominator is divisible by k, the numerator must also be so divisible. Hence removing trailing zeroes *also* ensures that no denominator is divisible by k. Let M be a DFA of n states accepting L.

We first show that the set $[\pi_2(L)]_k$ of possible denominators represented by L is finite. Write $S = S_1 \cup S_2$, where $S_1 = S \cap I[0, k^{n+1})$ and $S_2 = S \cap I[k^{n+1}, \infty)$. Let $L_1 = L \cap L_{<k^{n+1}}$, the representations of all numbers $< k^{n+1}$ in L, and $L_2 = L \cap L_{\geq k^{n+1}}$. Both L_1 and L_2 are regular, by Lemma 14. It now suffices to show that S_2 is $(\mathbb{N}, k)$-automatic.

Consider any $t \in S_2$. Let $z \in L_2$ be a representation of t. Since $t \geq k^{n+1}$, clearly $|z| \geq n$, and so $\pi_2(z)$ must begin with at least n 0's. Then, by the pumping lemma, we can write $z = uvw$ with $|uv| \leq n$ and $|v| \geq 1$ such that $uv^i w \in L$ for all $i \geq 0$. However, by the previous remark about $\pi_2(z)$, we see that $\pi_2(v) = 0^j$ for $1 \leq j \leq n$. Hence $[\pi_2(z)]_k = [\pi_2(uvw)]_k = [\pi_2(uw)]_k$. Since uw must also represent a member of S, it must be an integer, and hence $[\pi_2(z)]_k \mid [\pi_1(uw)]_k$ as well as $[\pi_2(z)]_k \mid [\pi_1(uvw)]_k$. Hence

$$[\pi_2(z)]_k \mid [\pi_1(uvw)]_k - [\pi_1(uw)]_k = ([\pi_1(uv)]_k - [\pi_1(u)]_k) \cdot k^{|w|}.$$

The previous reasoning applies to any $z \in L_2$. Furthermore, $0 < [\pi_1(uv)]_k - [\pi_1(u)]_k < k^n$. It follows that every possible denominator d of elements in L_2 can be expressed as $d = d_1 \cdot d_2$, where $1 \le d_1 < k^n$ and $d_2 \mid k^m$ for some m. It follows that the set of primes dividing all denominators d is finite, and we can therefore apply Theorem 20. Since k divides no denominator, the set of possible denominators is finite.

We can therefore decompose L_2 into a finite disjoint union corresponding to each possible denominator d. Next, we use a finite-state transducer to divide the numerator and denominator of the corresponding representations by d. For each d, this gives a new regular language A_d where the denominator is 1. Writing $T := \bigcup_d A_d$, we have $S_2 = \mathrm{quo}_k(T) = \bigcup_d \mathrm{quo}_k(A_d)$. Now we project, throwing away the second coordinate of elements of T; the result is regular and hence S is a k-automatic set of integers. □

Corollary 21. *Let $L \subseteq (\Sigma_k^2)^*$ be a regular language of words with no leading or trailing $[0,0]$'s, and suppose $\mathrm{quo}_k(L) \subseteq \mathbb{N}$. Then L can (effectively) be expressed as the finite union*

$$\bigcup_{1 \le i \le A} \{(ma_i, m)_k \ : \ m \in S_i\} \ \cup \ \bigcup_{1 \le j \le B} \{(b_j n, b_j)_k \ : \ n \in T_j\},$$

where $A, B \ge 0$ are integers, and $S_1, S_2, \ldots, S_A$ and $T_1, T_2, \ldots T_B$ are $(\mathbb{N}, k)$-automatic sets of integers, and $a_1, a_2, \ldots, a_A$ and $b_1, b_2, \ldots, b_B$ are non-negative integers.

Proof. Following the proof of Theorem 17, we see that L can be written as the union of L_1, the representations of integers $< k^{n+1}$ and L_2, the representations of those $\ge k^{n+1}$. For each integer $a < k^{n+1}$ we can consider the words of L whose quotient gives a; this provides a partition of L_1 into regular subsets corresponding to each a, and gives the first term of the union. For L_2, the proof of Theorem 17 shows that there are only a finite number of possible denominators, and the sets of corresponding numerators are k-automatic. □

As corollaries, we get that the k-automatic sets of rationals are (in contrast with sets of integers) not necessarily closed under the operations of intersection and complement.

Theorem 22. *Let $S_1 = \{(k^n - 1)/(k^m - 1) \ : \ 1 \le m < n\}$ and $S_2 = \mathbb{N}$. Then S_1 and S_2 are both k-automatic sets of rationals, but $S_1 \cap S_2$ is not.*

Proof. We can write every element of S_1 as p/q, where $p = (k^n - 1)/(k - 1)$ and $q = (k^m - 1)/(k - 1)$. The base-k representation of p is 1^n and the base-k representation of q is 1^m. Thus a representation for S_1 is given by the regular expression $[1,0]^+[1,1]^+$. We know that $\mathbb{N}$ is k-automatic from Example 5.

From a classical result we know that $(k^m - 1) \mid (k^n - 1)$ if and only if $m \mid n$. It follows that $S_1 \cap S_2 = T$, where

$$T = \{(k^n - 1)/(k^m - 1) \ : \ 1 < m < n \text{ and } m \mid n\}.$$

If the k-automatic sets of rationals were closed under intersection, then T would be $(\mathbb{N}, k)$-automatic. Writing $n = md$, we have

$$\frac{k^n - 1}{k^m - 1} = k^{(d-1)m} + \cdots + k^m + 1,$$

whose base-k representation is $(10^{m-1})^{d-1}1$. Hence $(T)_k = \{(10^{m-1})^{d-1}1 \,:\, m \geq 1, d > 1\}$. Assume this is regular. Intersecting with the regular language $10^*10^*10^*1$ we get $\{10^n10^n10^n1 \,:\, n \geq 1\}$. But a routine argument using the pumping lemma shows this is not even context-free, a contradiction. $\square$

From this result we can obtain several corollaries of interest.

Corollary 23. *The class of $(\mathbb{Q}^{\geq 0}, k)$-automatic sets is not closed under the operations of intersection or complement.*

Proof. We have just shown that this class is not closed under intersection. But since it is closed under union, if it were closed under complement, too, it would be closed under intersection, a contradiction. $\square$

Corollary 24. *Define a "normalization operation" N that maps a word w to its canonical expansion in lowest terms $(p/d, q/d)_k$, where $d = \gcd(p, q)$ and $p = \pi_1(w)$, $q = \pi_2(w)$, and define $N(L) = \{N(w) \,:\, w \in L\}$. Then the operation N does not, in general, preserve regularity.*

Proof. We give an example of a regular language L where $N(L)$ is not even context-free. It suffices to take $L = (S_1)_k$, where S_1 is the set defined above in the statement of Theorem 22. Consider $N(L)$; then we know from the argument above that $S_1 \cap \mathbb{N} = T$, where

$$T = \{(k^n - 1)/(k^m - 1) \,:\, 1 < m < n \text{ and } m \mid n\}.$$

Hence $N(L) \cap (\mathbb{N})_k = (T)_k$, but from the argument above we know that $(T)_k = \{(10^{m-1})^{d-1}1 \,:\, m \geq 1, d > 1\}$ is not context-free. $\square$

Corollary 25. *There is a k-automatic set of rationals whose base-k expansions (as real numbers) are not accepted by any Büchi automaton.*

Proof. Define $S_3 = \{(k^m - 1)/(k^n - 1) \,:\, 1 \leq m < n\}$; this is easily seen to be a k-automatic set of rationals. However, the set of its base-k expansions is of the form

$$\bigcup_{0<m<n<\infty} 0.(0^{n-m}\,(k-1)^m)^\omega,$$

where by x^ω we mean the infinite word $xxx\cdots$. A simple argument using the pumping lemma shows that no Büchi automaton can accept this language. $\square$

It follows that the class of k-automatic sets of rational numbers we study in this paper is not the same as that studied in [1, 4, 5, 6].

Remark 26. The technique above also allows us to prove that the languages $L_d = \{(p, q)_k \,:\, q \mid p\}$, $L_r = \{(p, q)_k \,:\, \gcd(p, q) > 1\}$, and $L_g = \{(p, q)_k \,:\, \gcd(p, q) = 1\}$ are not context-free.

For the first, suppose L_d is context-free and is accepted by a PDA M_d. Consider a PDA M that on input a unary word $x := 1^n$, $n \geq 2$, guesses a word of the form $y := 0^{n-a}1^a$ with

$1 < a < n$ and simulates M_d on $x \times y$, accepting if and only if $x \times y \in L_d$. Then M accepts the unary language $\{1^n \ : \ n \text{ composite}\}$ which is well-known to be non-context-free [14, Ex. 6.1, p. 141], a contradiction.

Notice that L_d can be considered as the set of all possible rational representations of $\mathbb{N}$.

For L_r, the same kind of construction works.

For L_g, a more complicated argument is needed. First, we prove that the language $L_c = \{0^i 1^j \ : \ i, j \geq 1 \text{ and } \gcd(i,j) > 1\}$ is not context-free. To see this, assume it is, use the pumping lemma, let n be the constant, and let p be a prime $> n$. Choose $z = 0^{p^2} 1^p \in L_c$. Then we can write $z = uvwxy$ where $|vwx| \leq n$ and $|vx| \geq 1$ and $uv^i wx^i y \in L_c$ for all $i \geq 0$. Then v and x each contain only one type of letter, because otherwise uv^2wx^2y has 1's before 0's, a contradiction. There are three possibilities:

(a) $vx = 0^r$, $1 \leq r \leq n$;

(b) $vx = 1^s$, $1 \leq s \leq n$;

(c) $v = 0^r$ and $x = 0^s$, $1 \leq r, s \leq n$.

In case (a) we pump with $i = 2$, obtaining $uv^2xw^2y = 0^{p^2+r}1^p$. But p is a prime, so for this word to be in L_c we must have $r \equiv 0 \pmod{p}$, contradicting the inequality $1 \leq r \leq n < p$.

Similarly, in case (b) we get $0^{p^2}1^{p+s}$. But $1 \leq s < p$, so $p < p + s < 2p$. Hence p does not divide $p + s$, so $\gcd(p^2, p + s) = 1$, a contradiction.

Finally, in case (c) and pumping with $i = j + 1$ we get $0^{p^2+rj}1^{p+sj}$. Since p is a prime, and $1 \leq r, s < p$, by Dirichlet's theorem there are infinitely many primes of the form $p^2 + rj$ and $p + sj$. If $r \geq s$, choose j so that $p^2 + rj$ is a prime. Since $p^2 + rj > p + sj$, we have $\gcd(p^2 + rj, p + sj) = 1$, a contradiction.

If $r < s$, choose $j > p^2 - p$ such that $p + sj$ is prime. Then $p + j > p^2$, so by adding $(s-1)j$ to both sides we get $p + sj > p^2 + (s-1)j \geq p^2 + rj$. Thus $p + sj$ is a prime greater than $p^2 + rj$ and so $\gcd(p^2 + rj, p + sj) = 1$, a contradiction. This completes the proof.

From this we also get that

$$L_e = \{0^i 1^j \ : \ \gcd(i,j) = 1\}$$

is not context-free, since it is known that the class of context-free languages that are subsets of 0^*1^* are closed under relative complement with 0^*1^* [12].

We can now prove that $L_g = \{(p,q)_k \ : \ \gcd(p,q) = 1\}$ is not context-free. Suppose it is. Then, since the CFL's are closed under intersection with a regular language, $L' := L_g \bigcap ([1,0]^*[1,1]^+ \ \cup \ [0,1]^*[1,1]^+)$ is also context-free. But the numerators are numbers of the form $(k^n - 1)/(k-1)$ and the denominators are numbers of the form $(k^m - 1)/(k-1)$. From a classical result, we know that $\gcd(k^m - 1, k^n - 1) = k^{\gcd(m,n)} - 1$. So it follows that

$$L' = \{[1,0]^s[1,1]^t \ : \ \gcd(s+t,t) = 1 \text{ and } s \geq 0, t \geq 1\}$$
$$\cup \ \{[0,1]^s[1,1]^t \ : \ \gcd(s+t,t) = 1 \text{ and } s \geq 0, t \geq 1\}.$$

Now apply the morphism h that maps $[a, b]$ to $a + b - 1$. Since the CFL's are closed under morphism, we get

$$h(L') = \{0^m 1^n \ : \ \gcd(m, n) = 1 \text{ and } m, n \geq 1\}.$$

But we already proved this is not context-free, a contradiction. Hence L_g is not context-free.

5 Solvability results

In this section we show that a number of problems involving k-automatic sets of integers and rational numbers are recursively solvable.

Theorem 27. *The following problems are recursively solvable: given a DFA M, a rational number α, and a relation $\lhd$ chosen from $=, \neq, <, \leq, >, \geq$, does there exist $x \in \mathrm{quo}_k(L(M))$ with $x \lhd \alpha$?*

Proof. The following gives a procedure for deciding if $x \lhd \alpha$. First, we create a DFA M' accepting the language $L_{\lhd\alpha}$ as described in Lemma 14 above. Next, using the usual direct product construction, we create a DFA M'' accepting $L(M) \cap L_{\lhd\alpha}$. Then, using breadth-first or depth-first search, we check to see whether there exists a path from the initial state of M'' to some final state of M''. Since by definition $L_{\lhd\alpha}$ contains every representation of each rational $x \in \mathrm{quo}_k(L_{\lhd\alpha})$, we have $\mathrm{quo}_k(L(M) \cap L_{\lhd\alpha}) = \mathrm{quo}_k(L(M)) \cap \mathrm{quo}_k(L_{\lhd\alpha})$, and therefore this procedure is correct. □

Lemma 28. *Let M be a DFA with input alphabet Σ_k^2 and let $F \subseteq \mathbb{Q}^{\geq 0}$ be a finite set of non-negative rational numbers. Then the following problems are recursively solvable:*

1. *Is $F \subseteq \mathrm{quo}_k(L(M))$?*
2. *Is $\mathrm{quo}_k(L(M)) \subseteq F$?*

Proof. To decide if $F \subseteq \mathrm{quo}_k(L(M))$, we simply check, using Lemma 14, whether $x \in \mathrm{quo}_k(L(M))$ for each $x \in F$.

To decide if $\mathrm{quo}_k(L(M)) \subseteq F$, we create DFA's accepting $L_{=x}$ for each $x \in F$, using Lemma 14. Now we create an automaton accepting all representations of all elements of F using the usual direct product construction for the union of regular languages. Since F is finite, the resulting automaton A is finite. Now, using the usual direct product construction, we create a DFA accepting $L(M) - L(A)$ and check to see if its language is empty. □

Theorem 29. *The following problem is recursively solvable: given a DFA M, and an integer k, is the set $\mathrm{quo}_k(L(M))$ infinite?*

Note that this is *not* the same as asking whether the language $L(M)$ itself is infinite, since a single number may have infinitely many representations.

First, we need a useful lemma from [23, 22].

Lemma 30. *Let $u, v, w \in (\Sigma_k^2)^*$ such that $|v| \geq 1$, and such that $[\pi_1(uvw)]_k$ and $[\pi_2(uvw)]_k$ are not both* 0*. Define*

$$U := \begin{cases} \mathrm{quo}_k(w), & \text{if } [\pi_1(uv)]_k = [\pi_2(uv)]_k = 0; \\ \infty, & \text{if } [\pi_1(uv)]_k > 0 \text{ and } [\pi_2(uv)]_k = 0; \\ \frac{[\pi_1(uv)]_k - [\pi_1(u)]_k}{[\pi_2(uv)]_k - [\pi_2(u)]_k}, & \text{otherwise.} \end{cases} \tag{6}$$

(a) Then exactly one of the following cases occurs:

(i) $\mathrm{quo}_k(uw) < \mathrm{quo}_k(uvw) < \mathrm{quo}_k(uv^2w) < \cdots < U$ *;*

(ii) $\mathrm{quo}_k(uw) = \mathrm{quo}_k(uvw) = \mathrm{quo}_k(uv^2w) = \cdots = U$ *;*

(iii) $\mathrm{quo}_k(uw) > \mathrm{quo}_k(uvw) > \mathrm{quo}_k(uv^2w) > \cdots > U$ *.*

(b) Furthermore, $\lim_{i\to\infty} \mathrm{quo}_k(uv^iw) = U$*.*

Now we can prove Theorem 29.

Proof. Without loss of generality, we may assume that the representations in M are canonical (contain no leading $[0,0]$'s). Define

$$\gamma_k(u, v) = \frac{[\pi_1(uv)]_k - [\pi_1(u)]_k}{[\pi_2(uv)]_k - [\pi_2(u)]_k},$$

and let $\mathrm{pref}(L)$ denote the language of all prefixes of all words of L. Let n be the number of states in M. We claim that the set $\mathrm{quo}_k(L(M))$ is finite if and only if $\mathrm{quo}_k(L(M)) \subseteq T$, where

$$T = \{\mathrm{quo}_k(x) \,:\, x \in L(M) \text{ and } |x| < n\} \cup \{\gamma_k(u, v) \,:\, uv \in \mathrm{pref}(L) \text{ and } |v| \geq 1 \text{ and } |uv| \leq n\}. \tag{7}$$

One direction is easy, since if $\mathrm{quo}_k(L(M)) \subseteq T$, then clearly the set $\mathrm{quo}_k(L(M))$ is finite, since T is.

Now suppose $\mathrm{quo}_k(L(M)) \subsetneq T$, so there exists some $x \in L(M)$ with $\mathrm{quo}_k(x) \notin T$. Since T contains all words of $L(M)$ of length $< n$, such an x is of length $\geq n$. So the pumping lemma applies, and there exists a decomposition $x = uvw$ with $|uv| \leq n$ and $|v| \geq 1$ such that $uv^iw \in L$ for all $i \geq 0$. Now apply Lemma 30. If case (ii) of that lemma applies, then $\mathrm{quo}_k(x) = \gamma_k(u, v) \in T$, a contradiction. Hence either case (i) or case (iii) must apply, and the lemma shows that $\mathrm{quo}_k(uv^iw)$ for $i \geq 0$ gives infinitely many distinct elements of $\mathrm{quo}_k(L(M))$.

To solve the decision problem, we can now simply enumerate the elements of T and use Lemma 28. □

Theorem 31. *Given $p/q \in \mathbb{Q}^{\geq 0}$, and a DFA M accepting a k-automatic set of rationals S, it is decidable if p/q is an accumulation point of S.*

Proof. The number α is an accumulation point of a set of real numbers S if and only if at least one of the following two conditions holds:

(i) $\alpha = \sup(\, S \ \cap \ I(-\infty, \alpha)\,)$;

(ii) $\alpha = \inf(\, S \ \cap \ I(\alpha, \infty)\,)$.

By Lemma 14 we can compute a DFA accepting $S' := S \ \cap \ I(-\infty, \alpha)$ (resp., $S \ \cap \ I(\alpha, \infty)$). By [23, Thm. 2] we can compute $\sup S'$ (resp., $\inf S'$). $\square$

Theorem 32. *Suppose S is a k-automatic set of integers accepted by a finite automaton M. There is an algorithm to decide, given M, whether there exists a finite set $D \subseteq \mathcal{P}$ such that $S \subseteq \pi(D)$. Furthermore, if such a D exists, we can also determine the sets F and U_f in Theorem 20.*

Proof. To determine if such a D exists, it suffices to remove all trailing zeroes from words in $(S)_k$ and see if the resulting language is finite. If it is, we know F, and then it is a simple matter to compute the U_f. $\square$

Theorem 33. *There is an algorithm that, given a DFA M_1 accepting $L_1 \subseteq (\Sigma_k^2)^*$, will determine if $\mathrm{quo}_k(L_1) \subseteq \mathbb{N}$. If so, the algorithm produces an automaton M_2 such that $[L(M_2)]_k = \mathrm{quo}_k(L_1)$.*

Proof. We start with the algorithm to determine if $\mathrm{quo}_k(L_1) \subseteq \mathbb{N}$.

Modify M_1, if necessary, to accept only representations of nonzero numbers, and to accept the remaining words of L_1 stripped of leading and trailing $[0, 0]$'s. Let M_1 have n states.

1. Create, using Lemma 14, a DFA M_3 accepting a representation of the set $T_1 := (\mathrm{quo}_k(L_1) \ \cap \ I(0, k^{n+1})) \setminus \{0, 1, \ldots, k^{n+1} - 1\}$. If $T_1 \neq \emptyset$, then answer "no" and stop.

2. Next, create a DFA M_4 accepting the set $T_2 := (\mathrm{quo}_k(L_1) \ \cap \ I[k^{n+1}, \infty))$. If any denominator ends in 0, answer "no" and stop.

3. Compute $\pi_2(L(M_4))$ and, using Theorem 32, decide if the integers so represented are factorable into a finite set of primes. If not, answer "no" and stop.

4. Otherwise, compute the decomposition in Corollary 21, obtaining the finite set of denominators in that decomposition. Check whether each denominator divides all of the corresponding numerators. If not, answer "no" and stop. Otherwise, answer "yes".

To see that this works, note that if some non-integer α belongs to $\mathrm{quo}_k(L_1)$, then either $\alpha < k^{n+1}$ or $\alpha \geq k^{n+1}$. In the former case we have $T_1 \neq \emptyset$, so this will be detected in step 1.

Otherwise $\alpha \geq k^{n+1}$. Then either the set of denominators of L_1 do not factor into a finite set of primes (which will be detected in step 3), or they do. In the latter case the set of denominators is k-finite, by Theorem 20 and so has a representation in the form given by Lemma 18 (b). Now k cannot divide both a numerator and denominator, because we have removed trailing $[0, 0]$s from every representation. So k divides a numerator but not a denominator if and only if this is detected in step 2. If steps 2 and 3 succeed, then, there are only finitely many denominators.

Now $\mathrm{quo}_k(L_1) \subseteq \mathbb{N}$ if and only if the numerators n corresponding to each of these finitely many denominators d are actually divisible by d. We can (effectively) form a partition of

L_1 according to each denominator. Using a machine to test divisibility by d, we can then intersect with each corresponding machine in the partition to ensure each numerator is indeed divisible. If so, we can easily produce an $(\mathbb{N}, k)$-automaton accepting $\text{quo}_k(L_1)$. $\square$

Corollary 34. *There is an algorithm that, given a DFA M_1 accepting $L_1 \subseteq (\Sigma_k^2)^*$ and a DFA M_2 accepting $L_2 \subseteq (\Sigma_k)^*$, will decide*

(a) if $\text{quo}_k(L_1) \subseteq [L_2]_k$*;*

(b) if $\text{quo}_k(L_1) = [L_2]_k$*.*

Proof. (a) The algorithm is as follows: using the algorithm in the proof of Theorem 33, first determine if $\text{quo}_k(L_1) \subseteq \mathbb{N}$. If not, answer "no". If so, using the algorithm in that proof, we determine an automaton M such that $\text{quo}_k(L_1) = [L(M)]_k$. Finally, using the usual cross-product construction, we create an automaton M' that accepts $L(M) \setminus L_2$. If M' accepts anything, then answer "no"; otherwise answer "yes".

(b) Similar to the previous case. In the last step, we create an automaton M' that accepts the symmetric difference $(L(M) \setminus L_2) \bigcup (L_2 \setminus L(M))$. $\square$

In particular, it is decidable if $\text{quo}_k(L_1) = \mathbb{N}$.

6 Subword complexity of denominators

The *subword complexity* of a language $L \subseteq \Sigma^*$ is the function $f_L : \mathbb{N} \to \mathbb{N}$ defined by $f_L(n) = |\Sigma^n \cap L|$, the number of distinct words of length n in L.

The following is a natural question; suppose L is a language such that $\text{quo}_k(L) = \mathbb{Q}^{\geq 0}$. What is the smallest possible subword complexity of the denominators $\pi_2(L)$? If no further restrictions on L are given, then it is easy to see that $f_{\pi_2(L)}$ can grow arbitrarily slowly (say, by enumerating the rational numbers and then finding arbitrarily large representations for each one). However, if L is restricted to be regular, then the best we can do is quadratic, as the following two results show.

Theorem 35. *If L is a regular language such that* $\text{quo}_k(L) = \mathbb{Q}^{\geq 0}$*, then $\pi_2(L)$ (resp., $\pi_1(L)$) is not of subword complexity $o(n^2)$.*

Proof. We prove the result for π_2; an analogous proof works for π_1.

Suppose there exists a regular language L with $\pi_2(L) = o(n^2)$. Then by a theorem of Szilard et al. [24], we know that the subword complexity of $\pi_2(L)$ must be $O(n)$. By another theorem in that same paper, we know that this implies that we can write $\pi_2(L)$ as the finite union

$$\pi_2(L) = \bigcup_{1 \leq i \leq n} u_i v_i^* w_i x_i^* y_i$$

where the u_i, v_i, w_i, x_i, y_i are (possibly empty) finite words.

Suppose every $u_i v_i$, for $1 \leq i \leq n$, contains a nonzero symbol. Then for every word $z \in L$ we would have $\text{quo}_k(z) < k^M$, where $M = \max_{1 \leq i \leq n} |u_i v_i|$, and hence we could not represent arbitrarily large rational numbers, a contradiction. It follows that there must be some nonempty subset $\mathcal{S}$ of the indices $\{1, 2, \ldots, n\}$ such that $u_i v_i \in 0^*$ for all $i \in \mathcal{S}$.

Similarly, suppose every $x_i y_i$, for $i \in \mathcal{S}$, contains a nonzero symbol. Then for every word $z \in L$ with $\pi_2(z) \in \bigcup_{i \in \mathcal{S}} u_i v_i^* w_i x_i^* y_i$ we would have $\nu_k(\pi_2(z)) < N$, where $N = \max_{1 \leq i \leq n} |x_i y_i|$. But then we could not represent all rational numbers of the form $(kp + 1)/(qk^j)$, where $j > N$ and $kp + 1 > k^{M+j} q$, a contradiction. It follows that there must be some nonempty subset $\mathcal{S}'$ of $\mathcal{S}$ such that both $u_i v_i$ and $x_i y_i$ are in 0^* for all $i \in \mathcal{S}'$.

From the above argument, all rational numbers of the form $(kp + 1)/(qk^j)$ with $j > N$ and $kp + 1 > k^{M+j} q$ must be represented by $z \in L$ with $\pi_2(z) \in \bigcup_{i \in \mathcal{S}'} u_i v_i^* w_i x_i^* y_i$. Since $u_i v_i$ and $x_i y_i$ are in 0^* for each such term, it follows that the set $\mathcal{T}$ of all prime factors of denominators of all these words is finite (and consists of the prime factors of k and $\pi_2(w_i)$ for $i \in \mathcal{S}'$). Choose any prime $r \notin \mathcal{T}$ and consider the rational number $r'/(rk^N)$, where r' is any prime with $r' > rk^{M+N}$. Then this number has no representation, a contradiction. $\square$

Theorem 36. *For each $k \geq 2$, there exists a regular language L such that* $\mathrm{quo}_k(L) = \mathbb{Q}^{\geq 0}$, *and $f_{\pi_2(L)}(n) = \Theta(n^2)$.*

Proof. Let
$$L_2 = \{w \in (\Sigma_k^2)^* \ : \ \pi_2(w) \in 0^* 1^+ 0^*\}$$
be the language given above in Example 7. We claim that there are exactly $n(n+1)/2$ words of length n in $\pi_2(L_2)$. To see this, count words of the form $0^* 1^+ 0^*$ of length n. There is 1 word consisting of all 1's, 2 words with $n - 1$ consecutive 1's, 3 words with $n - 2$ consecutive 1's, and so forth, for a total of $1 + 2 + \cdots + n = n(n + 1)/2$ words. $\square$

7 Open problems

There are a number of open problems raised by this work. The most outstanding one is a generalization of Cobham's theorem [8] to the setting of k-automatic sets of rationals.

Let $r \geq 1$ be an integer. We say a set $A \subseteq \mathbb{N}^r$ is *linear* if there exist vectors $v_0, v_1, \ldots, v_i \in \mathbb{N}^r$ such that
$$A = \{v_0 + a_1 v_1 + \cdots + a_i v_i \ : \ a_1, a_2, \ldots, a_i \in \mathbb{N}\}.$$
We say a set is *semilinear* if it is the finite union of linear sets.

Given a subset $A \subseteq \mathbb{N}^2$, we can define its set of quotients $q(A)$ to be $\{p/q \ : \ [p, q] \in A\}$.

Conjecture 37. *S is a set of rational numbers that is simultaneously k- and l-automatic for multiplicatively independent integers $k, l \geq 2$ if and only if there exists a semilinear set $A \subseteq \mathbb{N}^2$ such that $S = q(A)$.*

One direction of this conjecture is clear, as given A we can easily build an automaton to accept the base-k representation of the set of rationals $q(A)$. The converse, however, is not so clear.

We now turn to some decision problems whose status we have not been able to resolve. Which of the following problems, if any, are recursively solvable?

Given a DFA accepting $L \subseteq (\Sigma_k^2)^*$ representing a k-automatic set of rationals S,

1. does S contain at least one integer? (alternatively: is there $(p, q)_k \in L$ such that $q \mid p$?)
2. does S contain infinitely many integers?
3. are there infinitely many $(p, q)_k \in L$ such that $q \mid p$?
4. is there some rational number $p/q \in S$ having infinitely many distinct representations in L?
5. are there infinitely many distinct rational numbers $p/q \in S$ having infinitely many distinct representations in L?
6. do all rational numbers $p/q \in S$ have infinitely many distinct representations in L?
7. do all the $(p, q)_k \in L$ have $\gcd(p, q) = 1$?
8. does any $(p, q)_k \in L$ have $\gcd(p, q) = 1$?
9. is $\max_{(p,q)_k \in L} \gcd(p, q)$ bounded?

In the particular case where $S = \mathbb{Q}^{\geq 0}$, we have the following conjecture.

Conjecture 38. If L is a regular language with $\text{quo}_k(L) = \mathbb{Q}^{\geq 0}$, then L contains infinitely many distinct representations for infinitely many distinct rational numbers.

We note that it is possible to construct a regular L with $\text{quo}_k(L) = \mathbb{Q}^{\geq 0}$ where there are infinitely many rational numbers with exactly one representation. For example, if $k = 2$, we can take the language L_3 of Example 8 and add back a single representation for each element of the set $\{2^n \ : \ n \in \mathbb{Z}\}$.

Here are some additional decision problems whose status (recursively solvable or unsolvable) we have not been able to resolve:

Given DFA's M_1 and M_2 accepting languages $L_1, L_2 \subseteq (\Sigma_k^2)^*$ representing sets $S_1, S_2 \subseteq \mathbb{Q}^{\geq 0}$,

1. is $S_1 = S_2$?
2. is $S_1 \subseteq S_2$?
3. is $S_1 \cap S_2 \neq \emptyset$?

We hope to address some of these questions in a future paper.

It has long been known that $\text{Th}(\mathbb{N}, +, |)$, where $|$ denotes the divisibility relation, is undecidable; see, for example, [25]. We observe that the decidability of assertions involving even just two or three quantifiers, divisibility, and automata, would allow the solution of two classic open problems from number theory.

Example 39. Consider the language $L_7 \subseteq (\Sigma_2^2)^*$ defined by words with first component representing a numerator n of the form $2^i + 1$ for $i > 32$ and denominator an odd number d with $1 < d < n$. This is clearly regular. Now consider the assertion

$$\exists p \, \forall q \; (q \mid p) \implies (p,q)_2 \notin L_7.$$

This assertion is true if and only if there exists a Fermat prime greater than $2^{32} + 1$.

Example 40. Consider the language $L_8 \subseteq (\Sigma_2^2)^*$ defined by words with first component representing a numerator n of the form $2^i - 1$ for $i \geq 3$, and denominator an odd number d with $1 < d < n$. This is easily seen to be regular. Now consider the assertion

$$\forall t \; \exists p > t \; \forall q \; (q \mid p) \implies (p,q)_2 \notin L_8.$$

This assertion is true if and only if there are infinitely many Mersenne primes.

8 Acknowledgments

We thank Émilie Charlier, Luke Schaeffer, Jason Bell, Narad Rampersad, Jean-Paul Allouche, and Thomas Stoll for their helpful comments.

References

[1] Adamczewski, B., Bell, J.: An analogue of Cobham's theorem for fractals. Trans. Amer. Math. Soc. 363, 4421–4442 (2011)

[2] Allouche, J.-P., Shallit, J.: Automatic Sequences: Theory, Applications, Generalizations. Cambridge University Press (2003)

[3] Bés, A.: Undecidable extensions of Büchi arithmetic and Cobham-Semënov theorem, J. Symbolic Logic 62, 1280–1296 (1997)

[4] Boigelot, B., Brusten, J., Bruyère, V.: On the sets of real numbers recognized by finite automata in multiple bases. Log. Methods Comput. Sci. 6, no. 1, 1–17 (2010)

[5] Boigelot, B., Brusten, J.: A generalization of Cobham's theorem to automata over real numbers. Theoret. Comput. Sci. 410 no. 18, 1694–1703 (2009)

[6] Boigelot, B., Brusten, J., Leroux, J.: A generalization of Semenov's theorem to automata over real numbers. Automated deduction — CADE-22, 469–484, Lecture Notes in Comput. Sci., 5663, Springer, Berlin, (2009)

[7] Cantor, G.: Ueber unendliche, lineare Punktmannigfaltigkeiten, 5. Math. Annalen 21, 545–591 (1883)

[8] Cobham, A.: On the base-dependence of sets of numbers recognizable by finite automata. Math. Systems Theory 3, 186–192 (1969)

[9] Cobham, A.: Uniform tag sequences. Math. Systems Theory 6, 164–192 (1972)

[10] Eilenberg, S.: Automata, Languages, and Machines, Vol. A, Academic Press, 1974.

[11] Even, S.: Rational numbers and regular events. IEEE Trans. Electr. Comput. 13, 740–741 (1964)

[12] Ginsburg, S.: The Mathematical Theory of Context-Free Languages, McGraw-Hill (1966)

[13] Hartmanis, J., Stearns, R.E.: Sets of numbers defined by finite automata. Amer. Math. Monthly 74, 539–542 (1967)

[14] Hopcroft, J.E., Ullman, J.D.: Introduction to Automata Theory, Languages, and Computation, Addison-Wesley (1979)

[15] Loxton, J.H., van der Poorten, A.J.: An awful problem about integers in base four. Acta Arith 49, 193–203 (1987)

[16] Matos, A. B.: Periodic sets of integers. Theoret. Comput. Sci. 127, 287–312 (1994).

[17] Rowland, E., Shallit, J.: k-automatic sets of rational numbers. In: Dediu, A.-H., Martín-Vide, C. (eds.) Proceedings 6th International Conference, Language and Automata Theory and Applications 2012, pp. 490–501 (2012), vol. 7183 of *Lecture Notes in Computer Science*

[18] Salon, O.: Suites automatiques à multi-indices. Séminaire de Théorie des Nombres de Bordeaux 4(27), 4.01–4.27 (1986–1987)

[19] Salon, O.: Suites automatiques à multi-indices et algébricité. C. R. Acad. Sci. Paris 305, 501–504 (1987)

[20] Salon, O.: Propriétés arithmétiques des automates multidimensionnels. Thèse, Université Bordeaux I (1989)

[21] Schmid, J.: The joint distribution of the binary digits of integer multiples. Acta Arithmetica 43, 391–415 (1984)

[22] Schaeffer, L., Shallit, J.: The critical exponent is computable for automatic sequences. Int. J. Found. Comput. Sci. 23, 1611–1626 (2012)

[23] Shallit, J.: The critical exponent is computable for automatic sequences. In: Ambroz, P., Holub, S., Masáková, Z. (eds.) Proceedings 8th International Conference, WORDS 2011, pp. 231–239 (2011), vol. 63 of *Elect. Proc. Theor. Comput. Sci.*.

[24] Szilard, A., Yu, S., Zhang, K., Shallit, J.: Characterizing regular languages with polynomial densities, in I. M. Havel, V. Koubek, eds., Mathematical Foundations of Computer Science, 17th Intern. Symp., MFCS 1992, Lect. Notes in Comp. Sci., Vol. 629, pp. 494–503 (1992)

[25] Tarski, A.: Undecidable Theories. North-Holland (1953).