The power of one-letter rational languages ⋆

Thierry Cachat

Lehrstuhl für Informatik VII, RWTH, D-52056 Aachen
Fax: (49) 241-80-22215, Email: cachat@informatik.rwth-aachen.de

Abstract For any language L, let $\mathrm{pow}(L) = \{u^j \mid j \geqslant 0,\ u \in L\}$ be the set of powers of elements of L. Given a rational language L (over a finite alphabet), we study the question, posed by Calbrix 1996, whether $\mathrm{pow}(L)$ *is rational or not.* While leaving open the problem in general, we provide an algorithmic solution for the case of one-letter alphabets. This case is still non trivial; our solution is based on Dirichlet's result that for two relatively prime numbers, their associated arithmetic progression contains infinitely many primes.

1 Introduction

A great amount of work was done in studying rational languages and finite automata since the origins of the theory in 1956 (see [Kle 56] and [Per 95]). This theory is well developed and a lot of results are applicable via the effective properties of rational languages (see [HU 79] or [Per 90] for instance). The problem studied in this paper is easy to enunciate: we note

$$\mathrm{pow}(L) = \{u^j \mid j \geqslant 0,\ u \in L\} = \bigcup_{u \in L} u^* \quad \subseteq L^* \ ,$$

and consider the decision problem whether this language is rational if L is rational. More precisely, we are searching under which condition on L, $\mathrm{pow}(L)$ remains rational. This problem is far from trivial, it was first mentioned and left open in [Cal 96] (see also [CN 96]). As a simple example, consider the language $L \in Rat(\{a, b\}^*)$ defined by the rational expression ab^+. Its power $\mathrm{pow}(L) = \bigcup_{k>0} (ab^k)^*$ is context-sensitive but it is not context-free and of course not rational.

We recall the background of [Cal 96]. The rational $\omega-$languages are characterized by their ultimately periodic words, of the form uv^ω. Of course $uv^\omega = u(v^k)^\omega = uv^l(v^k)^\omega,\quad l, k \geqslant 0$. Given $M \subseteq A^\omega$ one defines its "periods":

$$\mathrm{per}(M) = \{v \in A^+ \mid \exists u \in A^*,\ uv^\omega \in M\} \subseteq A^* \ .$$

An important result is that if M is rational, then $\mathrm{per}(M)$ is rational. Note that $\mathrm{pow}(\mathrm{per}(M)) = \mathrm{per}(M)$. We can also consider a partial representation $L \in \mathrm{Rat}(A^*)$ such that $\mathrm{pow}(L) = \mathrm{per}(M)$, and ask whether $\mathrm{pow}(L) \in \mathrm{Rat}(A^*)$.

⋆ A previous version of this paper appeared in the Proceedings of the 5th international conference Developments in Language Theory, volume 2295 of Lecture Notes in Computer Science, pages 145–154. Springer, 2002.

For other classes of the Chomsky hierarchy than the regular languages, the question of the stability by the operation of power is easy and has a positive answer. We note the following facts:
- The power of a recursively enumerable language $(u_i)_{i \geqslant 0}$ is recursively enumerable too: one can enumerate $\left((u_i)^j\right)_{i,j \geqslant 0}$, the same way as $\mathbb{N}^2$.
- The power of a recursive language L is also recursive: a Turing Machine can, in a finite time, look for all decompositions of a word $u \in A^*$, in the form of a power $u = v^k$, and test if $v \in L$ (on another part of the band).
- We have the same result for a context-sensitive language.

In this paper, we restrict ourselves essentially to the special case of a one-letter alphabet (say $A = \{a\}$), and give an effective solution in Theorem 7. With this restriction, the languages can be easily represented as sets of integers, and we can use some elementary facts of arithmetic. In Section 3, we expose our main theorem and prove it. After that we mention possible future work in the conclusion.

2 Rational sets of integers

In this section we recall some classical results to handle rational sets of integers and give examples.

We first recall the notion of rational languages over a *finite* alphabet A (see [Per 90]): one denote $\mathrm{Rat}(A^*)$ the smallest set of languages which contains the finite languages and is closed under union ($\cup$), concatenation (.), and star (*) (the reflexive transitive closure of the concatenation). Since [Kle 56], it is known that $\mathrm{Rat}(A^*)$ is also the set of languages which are recognizable by finite automata, and is closed under complementation and intersection.

From now on we are considering a one-letter alphabet $A = \{a\}$ and remark that each word a^k of a^* is characterized by its length $k \in \mathbb{N}$. We identify a given $L \subseteq a^*$ with the set $\{k \geqslant 0 \mid a^k \in L\} \subseteq \mathbb{N}$. The structure $(a^*, \cup, \cdot, ^*)$ is isomorphic to $(\mathbb{N}, \cup, +, ^{\circledast})$, where $\cup$ is the usual union, $M + N = \{m + n \mid m \in M, n \in N\}$, and $M^{\circledast} = \{0\} \cup M \cup (M + M) \cup \cdots$. The set of rational languages of integers is denoted by $\mathrm{Rat}(\mathbb{N})$:

$$M \in \mathrm{Rat}(\mathbb{N}) \quad \Leftrightarrow \quad \{a^k \mid k \in M\} \in \mathrm{Rat}(a^*) \ .$$

Note that over a one-letter alphabet, the context-free languages are exactly the rational languages. According to [Per 90, p.36], $\mathrm{Rat}(\mathbb{N})$ is the set of 1-recognizable languages of integers: their representation in basis 1 is rational.
The product of sets of integers is defined with the usual multiplication:

$$\forall M, N \subseteq \mathbb{N}, \quad M.N = \{m.n \mid m \in M, n \in N\} \ .$$

Note that the operations + and . on sets of integers are still associative and commutative. We might omit the symbol (.) for the multiplication. For any

$L \subseteq a^*$, pow(L) is isomorphic to $M\mathbb{N}$, with $M = \{k \geqslant 0 \mid a^k \in L\}$, just because $(a^k)^j = a^{kj}$ for $j \geqslant 0$. That is to say, multiplication over the integers corresponds to the power operation over words. Now we can formulate our main question in terms of integers: given $M \in \mathrm{Rat}(\mathbb{N})$ we want to determine whether

$$M\mathbb{N} \in \mathrm{Rat}(\mathbb{N}) \ .$$

The previous remarks allow to characterize Rat$(\mathbb{N})$ directly in an intuitive way.

Lemma 1 (rational [Bü 59] [ES 69]) *For any $M \subseteq \mathbb{N}$, $M \in \mathrm{Rat}(\mathbb{N})$ iff M is ultimately periodic:*

$$\exists n > 0,\ \exists m \in \mathbb{N},\ I \subseteq [0, m),\ P \subseteq [m, m+n),\ M = I \cup (P + n\mathbb{N}) \ .$$

By convention $n\mathbb{N} = \{n\}\mathbb{N} = \{nk \mid k \in \mathbb{N}\}$, and $[a, b)$ is the segment of integers between a and b (a included, b excluded). In the situation of this lemma, we will say that n is eligible as a period for M, and (n, m, I, P) is *a representation* of the rational language M. Let $[x]_n = (x + n\mathbb{Z}) \cap \mathbb{N}$ be the class of x modulo n in $\mathbb{N}$. The rationality of M is also equivalent to the following condition:

$$\exists n > 0,\ \forall x \in [0, n),\ \ [x]_n \cap M \text{ is finite or co-finite.}$$

Example 2 *The language $0 \cup (5 + 2\mathbb{N})$ is rational, represented by* $(2, 5, \{0\}, \{5\})$..

Note that for languages of integers, the well known pumping lemma becomes an equivalence:

$$\begin{aligned} L \in \mathrm{Rat}(\mathbb{N}) &\Leftrightarrow \exists\ b \in \mathbb{N}, \forall\ l \in L, l \geqslant b, \exists\ 0 < r \leqslant b,\ l + r\mathbb{N} \subseteq L \\ &\Leftrightarrow \exists\ b \in \mathbb{N}, \exists\ 0 < r \leqslant b, \forall\ l \in L, l \geqslant b,\ l + r\mathbb{N} \subseteq L \ . \end{aligned}$$

This periodicity indicates that the "density" of a rational set is constant after a certain point, but we cannot always calculate it easily. The following example shows that it is not a sufficient condition of rationality.

Example 3 *Consider $M = \big((1 + 2\mathbb{N}) \cup \{2^k \mid k \geqslant 1\}\big) \setminus \{2^k + 1 \mid k \geqslant 1\}$. In each segment $[2m, 2m + 2)$ there is exactly one element of M, so we can say that the "density" is constant. But M is not ultimately periodic: its period would be greater than 2^k, for each $k \geqslant 1$.*

We still have a proposition that allows to prove (by contraposition) that some sets of the form $M\mathbb{N}$ are not rational. We note $|M|$ the cardinal of a (finite) set $M \subseteq \mathbb{N}$.

Proposition 4 *For all $L \in \mathrm{Rat}(\mathbb{N})$,*

$$\lim_{t\to\infty} \frac{|L\ \cap\ [0, t)|}{t} \text{ is well defined, and if this limit is 0, then } L \text{ is finite.}$$

Consider some simple examples of the product of a rational set by $\mathbb{N}$:

$$\begin{array}{ll} \mathbb{N}\mathbb{N} = (1+\mathbb{N})\mathbb{N} = (1+2\mathbb{N})\mathbb{N} = \mathbb{N} & \in \mathrm{Rat}(\mathbb{N}),\ \text{because } \mathbb{N}\mathbb{N} \supseteq 1.\mathbb{N} = \mathbb{N} \\ (2+2\mathbb{N})\mathbb{N} = 2(1+\mathbb{N})\mathbb{N} = 2\mathbb{N} & \in \mathrm{Rat}(\mathbb{N}),\ \text{but} \\ (3+2\mathbb{N})\mathbb{N} = \mathbb{N} \setminus 2^{\mathbb{N}} & \notin \mathrm{Rat}(\mathbb{N}). \end{array}$$

Where $2^{\mathbb{N}}$ is $\{2^k \mid k \in \mathbb{N}\}$. Indeed the set $(3+2\mathbb{N})\mathbb{N}$ consists of products of odd numbers greater than 1 and integers, so they cannot be a power of 2; conversely, an integer that is not a power of 2 has at least one odd prime factor. And the language $2^{\mathbb{N}}$ is not (ultimately) periodic.

To compute $M\mathbb{N}$ for a given M, a natural approach is to calculate $M \cup 2M \cup 3M \cdots$. Possibly that after some steps the union remains the same. In this case $M\mathbb{N}$ is rational as a finite union of rationals. In the other case this may not be true. For example $M = 2 \cup (3+2\mathbb{N})$ has the property that $M\mathbb{N} = \mathbb{N} \setminus \{1\} \in \mathrm{Rat}(\mathbb{N})$ (see Theorem 7), but the previous computation would be infinite (in brief $2^{n+1} \notin M.[0, 2^n)$).

In many cases we need the notions of division and prime number. The main idea of this paper is to use the basic properties of relative primality We write "$a \mid b$" for "a divides b". We assume that $a \mid 0$, even if $a = 0$. The set of prime numbers is denoted by $\mathscr{P}$.

We recall some results of arithmetic. They can be found e.g. in [AF 87] or [GKP 94] (Chap.4). The main fact we have used is Theorem 6 below. It is well known that $\mathscr{P}$ is infinite and not rational. The idea of the classical proof will be reused in Lemma 8. In fact there are less and less prime numbers along the integers (see [Ser 70] or [GKP 94]), but the density is decreasing very slowly.

Proposition 5 *For all $a, b \in \mathbb{N}$,*

$$a\mathbb{N} \cap b\mathbb{N} = \mathrm{lcm}(a,b)\mathbb{N} \qquad \text{and} \qquad ab = \mathrm{lcm}(a,b)\gcd(a,b)\ .$$

Theorem 6 (Dirichlet 1840) *For all $a, b > 0$,*

$$\gcd(a,b) = 1 \quad \Leftrightarrow \quad |\mathscr{P} \cap (a + b\mathbb{N})| = \infty\ .$$

The proof can be found in [HW 38] or [Ser 70]. We just give the easy direction: let $a, b > 0$. If $\gcd(a,b) = d$ and $d > 1$ then all the integers of $(a + b\mathbb{N})$ are divisible by d, so they contain at most one prime number, namely d.

3 The power of one-letter languages

In this section we will show our main result.

Theorem 7 *For a given language $L \in \mathrm{Rat}(\mathbb{N})$, one can decide algorithmically whether $L\mathbb{N} \in \mathrm{Rat}(\mathbb{N})$.*

We denote

$$Inv(m,q) = \{x \in [m, m+q) \mid \gcd(x,q) = 1\}\ ,$$

the set of integers relatively prime to q ("invertible" in $\mathbb{Z}/q\mathbb{Z}$), between m and $m+q-1$. To prove the theorem, we propose the following algorithm.

Algorithm: Rationality test for $L\mathbb{N}$, L rational

Input: $L \in \mathrm{Rat}(\mathbb{N})$ represented by (q, m, I, P) (see Lemma 1) with
$L = I \cup (P + q\mathbb{N})$, where $q \geqslant 1,\ m \geqslant 0,\ I \subseteq [0, m),\ P \subseteq [m, m+q)$
Output: "$L\mathbb{N} \in \mathrm{Rat}(\mathbb{N})$" or "$L\mathbb{N} \notin \mathrm{Rat}(\mathbb{N})$"

1 If $1 \in L$, then $L\mathbb{N} \in \mathrm{Rat}(\mathbb{N})$, end.
2 Else, if $\emptyset \neq Inv(m, q) \cap P \neq Inv(m, q)$, then $L\mathbb{N} \notin \mathrm{Rat}(\mathbb{N})$, end.
3 Else, if $Inv(m, q) \subseteq P$, obtain the answer with the equivalence:

$$L\mathbb{N} \in \mathrm{Rat}(\mathbb{N}) \quad \Leftrightarrow \quad \forall p \leqslant m + q \left(p \in \mathscr{P} \Rightarrow \exists b \geqslant 1,\ p^b \in L\right) \ .$$

end.
4 Else we have $\emptyset = Inv(m, q) \cap P$.
Compute, for each prime divisor u of q,

$$I_u = \left\{ \frac{x}{\gcd(u, x)} \;\middle|\; x \in I \right\} \ ,$$

$$P_u = \{x \in P \mid u \nmid x\} \cup \bigcup_{x \in P,\ u \mid x} \left\{ \frac{x}{u}, \frac{x+q}{u}, \cdots, \frac{x + q(u-1)}{u} \right\} \ ,$$

and $I'_u = (I_u \cup (P_u + q\mathbb{N})) \cap [0, m),\ P'_u = (P_u + q\mathbb{N}) \cap [m, m+q)$.
Let $L'_u = I'_u \cup (P'_u + q\mathbb{N})$. If $L'_u \neq L$ (*i.e.*, $I'_u \neq I$ or $P'_u \neq P$), then call recursively the algorithm with (q, m, I'_u, P'_u), to determine whether $L'_u\mathbb{N} \in \mathrm{Rat}(\mathbb{N})$. Collect every answer. Then answer with:

$$L\mathbb{N} \in \mathrm{Rat}(\mathbb{N}) \ \Leftrightarrow\ \forall u \in \mathscr{P}\, (u \mid q, L'_u \neq L \Rightarrow L'_u\mathbb{N} \in \mathrm{Rat}(\mathbb{N})) \ .$$

As a preparation to the correctness proof, we show the two following lemmas related to points 2 and 3 of the algorithm.

Lemma 8 *If $1 \notin L$ and $\emptyset \neq Inv(m, q) \cap P \neq Inv(m, q)$, then $L\mathbb{N} \notin \mathrm{Rat}(\mathbb{N})$.*

Proof: By hypothesis, we can choose $k \in Inv(m, q) \setminus P$. We know by Dirichlet's Theorem (6) that $k + q\mathbb{N}$ contains infinitely many prime numbers ($\gcd(k, q) = 1$). They are not in $L = I \cup (P + q\mathbb{N})$, because $k \notin P$. Let $p \in k + q\mathbb{N}$ be a prime number, $p \notin L$, and by hypothesis $1 \notin L$, so $p \notin L\mathbb{N}$: the only way to write p as a product is $1.p$ or $p.1$. As a consequence, $M = [k]_q \setminus L\mathbb{N}$ is infinite. We will see that it is not rational. By hypothesis, we also have some $s \in Inv(m, q) \cap P$. Let $n > 0$ and consider

$$f = s(s+q) \cdots (s+qn) \ .$$

The number f is relatively prime to q since s is, and so it is invertible in $\mathbb{Z}/q\mathbb{Z}$. Let $x > 0$ such that $x.f \equiv k - s \pmod q$. Remark that the integers $xf + s, xf + s + q, \cdots, xf + s + q.n$ are equivalent to k modulo q. But they are respectively divisible by $s, s+q, \cdots, s+qn$, since f is. So the numbers

$$\frac{xf + s}{s}, \frac{xf + s + q}{s + q}, \cdots, \frac{xf + s + qn}{s + qn}$$

are integers. Of course $s, s+q, \cdots, s+qn$ are in $(s+q\mathbb{N}) \subseteq L$, and so $xf+s, xf+s+q, \cdots, xf+s+qn$ are in $L\mathbb{N}$. As a consequence, the segment $[xf+s, xf+s+q(n+1))$ has an empty intersection with M.
This argument applies to any $n > 0$, and M infinite, so M is not ultimately periodic, hence $M \notin \mathrm{Rat}(\mathbb{N})$. We conclude that $L\mathbb{N}$ is not rational. ■

Lemma 9 *If $1 \notin L$ and $Inv(m,q) \subseteq P$, then*

$$L\mathbb{N} \in \mathrm{Rat}(\mathbb{N}) \quad \Leftrightarrow \quad \forall p \leqslant m+q \left(p \in \mathscr{P} \Rightarrow \exists b \geqslant 1, p^b \in L\right) \ .$$

Proof: If $Inv(m,q) \subseteq P$, then $P + q\mathbb{N}$ contains all numbers greater than m which are relatively prime to q. Thus it contains all prime numbers greater than $m+q$: if $p \in \mathscr{P}$, then p is relatively prime to q or p divides q; but if $p > q$, then $p \nmid q$. So only a finite number of primes is not in L. We denote $(p_k)_{k \geqslant 0}$ the *increasing* sequence of *all* prime numbers. Let $n \geqslant 0$ be such that $p_{n+1} > m+q$. The language L contains the p_k for $k > n$. One fixes $M = \mathbb{N} \setminus L\mathbb{N}$. For each $k \leqslant n$, let $b_k = \min(\{b \geqslant 0 \mid {p_k}^b \in L\} \cup \{\infty\})$. One can prove that

$$M \subseteq \{{p_0}^{a_0} \cdots {p_n}^{a_n} \mid a_0 < b_0, \cdots, a_n < b_n\} \ .$$

Indeed, for each $x \in \mathbb{N}$, if x has a prime factor $p_j, j > n$, then $p_j \in L$, and $x \in p_j\mathbb{N} \subseteq L\mathbb{N} \Rightarrow x \notin M$. So, if $x \in M$, then x is ${p_0}^{a_0} \cdots {p_n}^{a_n}$, but if there is a k such that $a_k \geqslant b_k$, then $x \in {p_k}^{b_k}\mathbb{N} \subseteq L\mathbb{N}$.

In the case that $\forall p \leqslant m+q,\ p \in \mathscr{P},\ \exists b \geqslant 1,\ p^b \in L$, each b_k is finite, and then M is finite (so rational), that is $L\mathbb{N}$ is co-finite, and rational. We conclude:

$$\forall p \leqslant m+q \left(p \in \mathscr{P} \Rightarrow \exists b \geqslant 1, p^b \in L\right) \quad \Rightarrow \quad L\mathbb{N} \in \mathrm{Rat}(\mathbb{N}) \ .$$

Conversely, in the case that $\exists j \leqslant n,\ \forall b \geqslant 1,\ {p_j}^b \notin L$, we have $b_j = \infty$, and M is infinite: it contains ${p_j}^{\mathbb{N}}$ ($1 \notin L$, and the only factors of ${p_j}^i$ are the ${p_j}^k, k \leqslant i$). We will prove that M is not rational with an argument of density inspired by [Cal 96]. We want to bound the number of elements of M lower than e^t, for $t \geqslant 1$:

$$\begin{aligned} &{p_0}^{a_0} \cdots {p_n}^{a_n} < e^t \qquad \Rightarrow \qquad a_0 \ln p_0 + \cdots + a_n \ln p_n < t \\ \Rightarrow \qquad &\forall k \leqslant n,\ a_k < \frac{t}{\ln p_k} \leqslant \frac{t}{\ln 2} \qquad \Rightarrow \qquad a_0, \cdots, a_n \in \left[0, \frac{t}{\ln 2}\right). \end{aligned}$$

That consists of at most $\left(\frac{t}{\ln 2} + 1\right)^{n+1}$ different (n+1)-tuples:

$$\left|M \cap [1, e^t)\right| < \left(\frac{t}{\ln 2} + 1\right)^{n+1} \quad \Rightarrow \quad \frac{|M \cap [1, e^t)|}{e^t} < \frac{\left(\frac{t}{\ln 2} + 1\right)^{n+1}}{e^t} \ \xrightarrow{t \to \infty} \ 0 \ .$$

The density of M has the limit zero, and thanks to Proposition 4 (M is infinite), we conclude that $M \notin \mathrm{Rat}(\mathbb{N})$, then $L\mathbb{N} \notin \mathrm{Rat}(\mathbb{N})$. ■

Now we prove the correctness of the algorithm for all cases.

Proof: (Theorem 7) We proceed in two steps: proof of partial correctness (under the assumption of termination), and proof of termination.

Proof of partial correctness
Let $L = I \cup (P + q\mathbb{N})$. We follow the algorithm step by step.

1- If $1 \in L$, then $L\mathbb{N} \supseteq 1\mathbb{N} = \mathbb{N}$, and $L\mathbb{N} = \mathbb{N} \in \mathrm{Rat}(\mathbb{N})$.
2- see Lemma 8
3- see Lemma 9
4- Else, we obtain $Inv(m, q) \cap P = \emptyset$. Each element of P has a (strict) common divisor with q, those of $P + q\mathbb{N}$ too, and those of $(P + q\mathbb{N})\mathbb{N}$ also.
The idea is to decompose the problem by considering the rationality of $u\mathbb{N} \cap L\mathbb{N}$ for some $u \in \mathscr{P}$. We have clearly

$$L\mathbb{N} \in \mathrm{Rat}(\mathbb{N}) \;\Rightarrow\; \forall u \in \mathscr{P}, u \mid q, (u\mathbb{N} \cap L\mathbb{N}) \in \mathrm{Rat}(\mathbb{N}) \ .$$

The problem is to find a kind of converse: to find a set $U \subseteq \mathscr{P} \cap [2, q]$ such that

$$L\mathbb{N} \in \mathrm{Rat}(\mathbb{N}) \;\Leftarrow\; \forall u \in U, (u\mathbb{N} \cap L\mathbb{N}) \in \mathrm{Rat}(\mathbb{N}) \ .$$

We claim that the set $U = \{u \in \mathscr{P} \mid u \mid q \text{ and } L'_u \neq L\}$ is convenient (and it is exactly what we need for the termination!). The set $U\mathbb{N}$ is rational as a finite union of rational sets, and

$$L\mathbb{N} \in \mathrm{Rat}(\mathbb{N}) \;\Leftrightarrow\; U\mathbb{N} \cap L\mathbb{N} \in \mathrm{Rat}(\mathbb{N}) \text{ and } (\mathbb{N} \setminus U\mathbb{N}) \cap L\mathbb{N} \in \mathrm{Rat}(\mathbb{N}) \ .$$

We will show later that $(P + q\mathbb{N})\mathbb{N} \subseteq U\mathbb{N}$. Assuming this fact we have $(\mathbb{N} \setminus U\mathbb{N}) \cap L\mathbb{N} = (\mathbb{N} \setminus U\mathbb{N}) \cap I\mathbb{N}$, which is rational ($I$ is finite). So we have

$$L\mathbb{N} \in \mathrm{Rat}(\mathbb{N}) \;\Leftrightarrow\; U\mathbb{N} \cap L\mathbb{N} \in \mathrm{Rat}(\mathbb{N}) \ .$$

We decompose into:

$$U\mathbb{N} \cap L\mathbb{N} = \left(\bigcup_{u \in U} u\mathbb{N}\right) \cap L\mathbb{N} = \bigcup_{u \in U} (u\mathbb{N} \cap L\mathbb{N}) \ ;$$

$$L\mathbb{N} \in \mathrm{Rat}(\mathbb{N}) \;\Leftrightarrow\; \forall u \in U \ (u\mathbb{N} \cap L\mathbb{N}) \in \mathrm{Rat}(\mathbb{N}) \ .$$

$$L\mathbb{N} \in \mathrm{Rat}(\mathbb{N}) \;\Leftrightarrow\; \forall u \in \mathscr{P} \, (u \mid q, L'_u \neq L \Rightarrow L'_u\mathbb{N} \in \mathrm{Rat}(\mathbb{N})) \ .$$

We still have to calculate the $u\mathbb{N} \cap L\mathbb{N}$, and then to show that $(P + q\mathbb{N})\mathbb{N} \subseteq U\mathbb{N}$, which proves that the set U is “convenient”.
Let $u \in \mathscr{P}, u \mid q$. For each $x \in P$, if $u \mid x$, then u divides each element of $x + q\mathbb{N}$, so $u\mathbb{N} \cap (x + q\mathbb{N})\mathbb{N} = (x + q\mathbb{N})\mathbb{N}$. Else $\gcd(u, x) = 1$ and u is also relatively prime to each integer of $x + q\mathbb{N}$, hence $u\mathbb{N} \cap (x + q\mathbb{N})\mathbb{N} = (x + q\mathbb{N})u\mathbb{N}$ (Prop. 5). In general

$$u\mathbb{N} \cap (P + q\mathbb{N})\mathbb{N} = \bigcup_{x \in P} (x + q\mathbb{N}) \frac{u}{\gcd(u, x)} \mathbb{N} \ .$$

Similarly for I,

$$u\mathbb{N} \cap I\mathbb{N} = \bigcup_{x \in I} \frac{xu}{\gcd(u, x)} \mathbb{N} \ .$$

The interest in calculating the intersection with $u\mathbb{N}$ is that all elements are divisible by u. Now:

$$\frac{u\mathbb{N} \cap L\mathbb{N}}{u} = \left(\bigcup_{x \in I} \frac{x}{\gcd(u,x)} \cup \bigcup_{x \in P} \left(\frac{x}{\gcd(u,x)} + \frac{q}{\gcd(u,x)} \mathbb{N} \right) \right) \mathbb{N} \ .$$

Consequently, $u\mathbb{N} \cap L\mathbb{N} \in \mathrm{Rat}(\mathbb{N})$ is equivalent to $L_u\mathbb{N} \in \mathrm{Rat}(\mathbb{N})$, with

$$L_u = I_u \cup (P_u + q\mathbb{N}) \ , \qquad I_u = \{x/\gcd(u,x) \mid x \in I\} \ ,$$

$$P_u = \{x \in P \mid u \nmid x\} \cup \bigcup_{x \in P,\ u|x} \left\{ \frac{x}{u}, \frac{x+q}{u}, \cdots, \frac{x+q(u-1)}{u} \right\} \ .$$

But this representation is not canonical with respect to Lemma 1, we must then consider:

$$I'_u = (I_u \cup (P_u + q\mathbb{N})) \cap [0,m) \ , \qquad P'_u = (P_u + q\mathbb{N}) \cap [m, m+q) \ ,$$

$$\text{and} \quad L'_u = I'_u \cup (P'_u + q\mathbb{N}) \ ,$$

as enunciated in the algorithm. One observes

$$L_u\mathbb{N} = L'_u\mathbb{N} = \frac{u\mathbb{N} \cap L\mathbb{N}}{u} \ .$$

Indeed, for each $x \in I$, x is replaced by x or x/u in L_u and L'_u. And for all $x \in P$, x is replaced by x or $x/u, (x+q)/u, \cdots, (x+q(u-1))/u$ in L_u. In the first case, x stay identically in L'_u, in the second each new element is smaller or equal than $(x+q(u-1))/u$, and

$$\frac{x+q(u-1)}{u} < \frac{m+q+q(u-1)}{u} \leqslant \frac{m}{2} + q \leqslant m+q \ ,$$

so the construction of L'_u does not omit any element, and $L_u = L'_u$. After that we can call the algorithm for (each) L'_u.

To prove that $(P+q\mathbb{N})\mathbb{N} \subseteq U\mathbb{N}$, it is sufficient to show that $\forall x \in P, \exists u \in U, u \mid x$, because

$$P \subseteq U\mathbb{N} \Rightarrow P + q\mathbb{N} \subseteq U\mathbb{N} \Rightarrow (P+q\mathbb{N})\mathbb{N} \subseteq U\mathbb{N}$$

(recall that $u \mid q$). Let $x \in P$. By hypothesis (case 4) $\exists u \in \mathscr{P}, u \mid x$ and $u \mid q$. The question is whether $L'_u \neq L$, *i.e.*, $u \in U$. We suppose by absurd that $\forall u \in \mathscr{P}, (u \mid x) \wedge (u \mid q) \Rightarrow L'_u = L$.
We consider a prime factor decomposition: $\gcd(x,q) = u_0 \cdots u_n$, where $\forall k \leqslant n, u_k \in \mathscr{P}$ (possibly $u_0 = u_1, \cdots$). By hypothesis $\forall k \leqslant n, L'_{u_k} = L$. Using the computation of the L'_u above, we see that

$$\frac{x}{u_0} + \frac{q}{u_0}\mathbb{N} \subseteq L_{u_0} = L \ .$$

But $u_1 \mid \frac{x}{u_0}$ and $u_1 \mid \frac{q}{u_0}$, so

$$\frac{x}{u_0 u_1} + \frac{q}{u_0 u_1}\mathbb{N} \subseteq L_{u_1} = L \ ,$$

and so on, by induction on n one can show

$$\frac{x}{u_0 \cdots u_n} + \frac{q}{u_0 \cdots u_n}\mathbb{N} \subseteq L \ .$$

Let $x' = \frac{x}{u_0 \cdots u_n}$, $q' = \frac{q}{u_0 \cdots u_n}$. By construction $\gcd(x', q') = 1$. By Theorem 6, $x' + q'\mathbb{N}$ contains infinitely many primes. In particular $\exists y \in (x' + q'\mathbb{N}) \cap \mathscr{P} \cap [m + q + 1, \infty)$, that is $y \in L \cap \mathscr{P}$ and $y > m + q$, so y is relatively prime to q, which is a contradiction with $\emptyset = Inv(m, q) \cap P$ (case 4).

So we conclude the partial correctness of the algorithm.

Proof of termination
Most of the steps of the algorithm are clearly effective, using the representation (q, m, I, P) of L. We just have to justify two points: there is a finite number of recursive calls, and we can determine if

$$\forall p \leqslant m + q, \ p \in \mathscr{P}, \ \exists b \geqslant 1, \ p^b \in L \ .$$

Let $p \in \mathscr{P}, \ p \leqslant m + q$. We compute the first powers of p: $p, p^2, \cdots, p^k$ with $p^k > m + q$. If one of them is in L, the condition is true for this p. Else, we can then calculate modulo q: $p^{k+1}, \cdots, p^{k+q} \pmod q$, since the following elements will not generate any new values modulo q. If one of them is in $P \pmod q$, the condition is true for p, else the condition is false (and $L\mathbb{N}$ is not rational).

For the recursive calls, we define the strict order $\prec$ on the (finite) sets of integers, by induction: $\forall A, B \subseteq \mathbb{N}, \ A, B \neq \emptyset$,

$$\begin{aligned} &\emptyset \prec B \\ &A \prec B \ \Leftrightarrow \ \begin{cases} \min(A) < \min(B), \ \text{or} \\ \min(A) = \min(B), \ \text{and } A \setminus \min(A) \prec B \setminus \min(B) \ . \end{cases} \end{aligned}$$

This is the lexicographical order over the "characteristic words" from $\{0, 1\}^\omega$ of A and B. The order $\prec$ is total.

We will prove that $I'_u \cup P'_u \prec I \cup P$, when $u \in U$. Let y be the smallest of the $x \in I \cup P$ such that $u \mid x$, there exists one because $L'_u \neq L$. By construction, $\forall x \in I \cup P, \ x < y \Rightarrow u \nmid x$, we find x also (identically) in $I'_u \cup P'_u$, and it does not "generate" any other element in $I'_u \cup P'_u$. On the other side y generates $y/u \in I'_u \cup P'_u$, and of course $y/u < y$, so $I'_u \cup P'_u \prec I \cup P$.

The integers m and q remain the same in each recursive call, and for fixed m and q, there is only a finite number (2^{m+q}) of possible sets $I \cup P \subseteq [0, m+q)$. So the order $\prec$, restricted to these sets, is finite. It follows that the computation of the algorithm for a given L, *i.e.*, (q, m, I, P), needs only (recursively) the computation of finitely many (q, m, I', P'), which prove the termination. ■

Example 10 *Considering $L = 3\mathbb{N} = \{0, 3, 6, \cdots\}$, one has $L\mathbb{N} = 3\mathbb{N}\mathbb{N} = 3\mathbb{N}$, so $L\mathbb{N} \in \mathrm{Rat}(\mathbb{N})$. One can represent L by $(3, 0, \emptyset, \{0\})$ and the algorithm would answer "rational" after a recursive call: $L'_3 = \mathbb{N}$. With the other representation $(6, 2, \{0\}, \{3, 6\}) = (q, m, I, P)$, i.e., $L = \{0\} \cup (\{3, 6\} + 6\mathbb{N})$, one sees that $u = 2$ satisfies $u \in \mathscr{P}$ and $u \mid q$, moreover $\exists x \in P, u \mid x$, but $L'_u = L$. Essentially because $\frac{L\cap 2\mathbb{N}}{2} = \frac{6\mathbb{N}}{2} = 3\mathbb{N}$. That is the reason why the condition $L'_u \neq L$ is necessary for the termination.*

To speed up the algorithm, it might be possible to factorize L by $\mathrm{lcm}(I \cup P \cup \{q\})$: the greatest common divisor of (all) the elements of $I \cup P \cup \{q\}$. One can also try to minimize the period q and the basis m.

As an application of the algorithm we can conclude that very often the power of rational languages is *not* rational. For example

$$\forall a, b \in \mathbb{N}, \quad (a + b\mathbb{N})\mathbb{N} \in \mathrm{Rat}(\mathbb{N}) \;\Leftrightarrow\; a \mid b \text{ or } b \mid a\ .$$

After some reductions of the languages through point 4 of the algorithm that can be done directly by a factorization with $\gcd(a, b)$, one obtains $a', b' \in \mathbb{N}, \gcd(a', b') = 1$. The cases a' or $b' \in \{0, 1\}$, corresponding to $a \mid b$ or $b \mid a$ are easy. Otherwise, if $b' = 2$, then $a' \geqslant 3$ and a' is odd, we conclude unrationality with point 3 and $\forall k \geqslant 0, 2^k \notin (a' + b'\mathbb{N})$. If $b' > 2$, point 2 shows that the language is unrational.

4 Conclusion

We have given an algorithmically solution to the question enunciated in Section 2:

$$\text{given } M \in \mathrm{Rat}(\mathbb{N}), \text{ is } M\mathbb{N} \text{ rational?}$$

Depending on the presence of relatively prime numbers in M, our algorithm either states directly that M is not rational using Dirichlet's theorem, or it considers the representation of all prime numbers in M, or it decomposes M into "smaller" languages by intersection with simple periodic sets and use recursivity. It was necessary to use some facts of arithmetic. The runtime is mainly influenced by the bound 2^{m+q} for the number of recursive calls; we do not give here a precise computation of the time complexity.

As future work, one could try to extend the result of Theorem 7 to any product of two rational languages of integers (or even not necessarily rational). It is easy to remark that if $L\mathbb{N} \notin \mathrm{Rat}(\mathbb{N})$ then $\forall t \geqslant 0,\ L(t + \mathbb{N}) \notin \mathrm{Rat}(\mathbb{N})$, but the converse proposition is false: $(2 + \mathbb{N})(2 + \mathbb{N}) = (4 + \mathbb{N}) \setminus \mathscr{P} \notin \mathrm{Rat}(\mathbb{N})$. In another direction, one can try to answer the question: "is $\mathrm{pow}(L)$ rational?" for any rational language L (over any finite alphabet). After that, the most general question would be: given a finite alphabet A, two languages $L \in \mathrm{Rat}(A^*)$ and $M \in \mathrm{Rat}(\mathbb{N})$,

$$\text{is } L^M = \{u^k \in A^* \mid u \in L, k \in M\} \text{ rational?}$$

5 Acknowledgment

Great thanks to Didier Caucal that gives us the subject of this paper, and a lot of advice, to Wolfgang Thomas and Tanguy Urvoy for their helpful comments and suggestions.

References

[AF 87] Jean-Marie ARNAUDIÈS and Henri FRAYSSE, *Cours de mathématiques - 1, Algèbre*, classes préparatoires, Bordas, 1987.

[Bü 59] J. Richard BÜCHI , *Weak second-order arithmetic and finite automata*, Zeit. für math. Logik und Grund. der Math. 6, pp. 66–92, 1960.

[Cal 96] Hugues CALBRIX, *Mots ultimement périodiques des langages rationnels de mots infinis*, Thèse de l'université Denis Diderot-Paris VII, Litp Th96/03, Avril 1996.

[CN 96] Hugues CALBRIX and Maurice NIVAT, Prefix and period languages of rational ω-languages, Proc. Developments in Language Theory, Magdeburg 1995, World Scientific, 1996, pp. 341–349.

[ES 69] Samuel EILENBERG and Marcel-Paul SCHÜTZENBERGER, Rational sets in commutative monoids, Journal of Algebra 13, pp. 173–191, 1969.

[GKP 94] Ronald L. GRAHAM, Donald E. KNUTH and Oren PATASHNIK, *Concrete mathematics*, Addison-Wesley, 1989, second edition 1994.

[HW 38] Godfrey H. HARDY and Edward M. WRIGHT, *An Introduction to the Theory of Numbers*, Clarendon Press, Oxford, 1938; fifth edition, 1979.

[HU 79] John E. HOPCROFT and Jeffrey D. ULLMAN, *Introduction to automata theory, languages and computation*, Addison-Wesley, 1979.

[Kle 56] Stephen C. KLEENE, *Representation of events in nerv nets and finite automata*, In C.E.Shannon and J.McCarthy, editors, Automata Studies, Princeton University Press, 1956.

[Mat 94] Armando B. MATOS, *Periodic sets of integers*, Theoretical Computer Science 127, Elsevier, 1994.

[Per 90] Dominique PERRIN, *Finite Automata*, Handbook of Theoretical Computer Science, J. van Leeuwen, Elsevier Science publishers B.V., 1990.

[Per 95] Dominique PERRIN, *Les débuts de la théorie des automates*, Technique et Science Informatique (vol. 14. 409–43), 1995.

[Ser 70] Jean-Pierre SERRE, *Cours d'arithmétique*, Presses Universitaires de France, 1970. Or *A course in arithmetic*, Springer-Verlag, New York-Heidelberg, 1973.