RWTH Aachen

Department of Computer Science
Technical Report

The power of one-letter rational languages

Thierry Cachat

ISSN 0935–3232 · Aachener Informatik Berichte · AIB-2001-03

RWTH Aachen · Department of Computer Science · March 2001

The power of one-letter rational languages

Thierry Cachat*

Lehrstuhl für Informatik 7, RWTH, D 52056 Aachen
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 in [Cal 96], 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]. As a simple example, consider the language $L \in Rat(\{a,b\}^*)$ defined by the rational expression ab^+. Its power $pow(L) = \bigcup_{k>0} (ab^k)^*$ is context-sensitive since context-sensitive languages are closed under power (see below) 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 $uv^\omega = u(v^k)^\omega = uv^l(v^k)^\omega,\ \ l,k \geqslant 0$. 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 $\mathrm{pow}(L) \in \mathrm{Rat}(A^*)$.

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

* This work was done at IRISA, Rennes, France in 2000.

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 31. With this restriction, the languages can be easily represented as sets of integers, and we can use some elementary facts of arithmetic (recalled in the Appendix). 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 $\mathrm{Rat}(\mathbb{N})$:

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

Remark 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 will omit the symbol $(*)$ for the multiplication. For any $L \subseteq a^*$, $\mathrm{pow}(L)$ 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 (Theorem 31) 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 $\mathrm{Rat}(\mathbb{N})$ directly in an intuitive way.

Lemma 21 (rational [Bü 59] [ES 69]) *For any $M \subseteq \mathbb{N}$, $M \in \mathrm{Rat}(\mathbb{N})$ iff M is ultimately periodic:*

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

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. 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:

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

Example 22 *The language $0 \cup (5 + 2\mathbb{N})$ is rational, with period 2 (or 4).*

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 23 *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 24 *For all $L \in \mathrm{Rat}(\mathbb{N})$,*

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

Proof: With Lemma 21, $L = (L \cap [0, m[) \cup ((L \cap [m, m + n[) + n\mathbb{N})$, and the limit is exactly $\frac{|[m, m+n[\ \cap\ L|}{n}$. ■

Let us now illustrate our main question above by some simple examples.

For any $M \subseteq \mathbb{N}$, 2^M is $\{2^k \mid k \in M\}$. We consider the product of a rational set by $\mathbb{N}$:

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

Indeed the set $(3+2\mathbb{N})\mathbb{N}$ consists of products of an odd number by an integer, 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 $M * 1 \cup M * 2 \cup M * 3 \cdots$. 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 we cannot conclude. 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 31), but the previous computation would be infinite (in brief $2^{n+1} \notin M.[0, 2^n[$).

In many examples it has turned out that we need the notions of division and prime number. The main idea of this paper is to use the basic properties of relative primality (see Appendix).

3 The power of one-letter languages

In this section we will show our main result.

Theorem 31 *For a given language $L \in \mathrm{Rat}(\mathbb{N})$, one can decide algorithmically whether $L\mathbb{N} \in \mathrm{Rat}(\mathbb{N})$.*

We note

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

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

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

Precondition: $q \geqslant 1,\ m \geqslant 0,\ I \subseteq [0, m[,\ P \subseteq [m, m+q[$
Input: $L \in \mathrm{Rat}(\mathbb{N})$ represented by (q, m, I, P) (see Lemma 21) with

$$L = I \cup (P + q\mathbb{N})$$

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:

$$L\mathbb{N} \in \mathrm{Rat}(\mathbb{N}) \quad \Leftrightarrow \quad \forall p \leqslant m+q,\ p \in \mathcal{P},\ \exists k \geqslant 1,\ p^k \in L,$$

end.

4 Else we have $\emptyset = Inv(m,q) \cap P$. Compute, for each prime divisor u of q such that $\exists x \in P,\ u \mid x$,

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

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

and $I'_u = (I_u \cup (P_u + q\mathbb{N})) \cap [0, m[,\ P'_u = (P_u + q\mathbb{N}) \cap [m, m+q[$.
Call recursively the algorithm with (q, m, I'_u, P'_u), to determine whether $L'_u\mathbb{N} \in \mathrm{Rat}(\mathbb{N})$, where $L'_u = I'_u \cup (P'_u + q\mathbb{N})$ and $L'_u\mathbb{N} = \frac{u\mathbb{N} \cap L\mathbb{N}}{u}$. Collect every answer. Then answer with:
$L\mathbb{N} \in \mathrm{Rat}(\mathbb{N}) \;\Leftrightarrow\; \forall u \in \mathcal{P}, u \mid q,$ such that $\exists x \in P, u \mid x, L'_u\mathbb{N} \in \mathrm{Rat}(\mathbb{N})$.

As a preparation to the correctness proof, we show the following Lemmas 32 and 33 related to points 2 and 3 of the algorithm.

Lemma 32 *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 (Appendix 52) 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 as well as s, 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 $k - s + s$, and to k, modulo q. But they are respectively divisible by $s, s+q, \cdots, s+qn$, as well as f. 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 can conclude that $L\mathbb{N}$ is not rational. ■

Lemma 33 *If $1 \notin L$ and $Inv(m,q) \subseteq P$, then*

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

Proof: If $Inv(m,q) \subseteq P$, then $P + q\mathbb{N}$ contains all numbers $> m$ which are relatively prime to q. Thus it contains all prime numbers greater than $m+q$: if $p \in \mathcal{P}$, 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 are not in L. We note $(p_k)_{k\geqslant 0}$ the *increasing* sequence of *all* prime numbers. Let $n \geqslant 0$ 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 $l_k = \min(\{l \geqslant 0 \mid {p_k}^l \in L\} \cup \{\infty\})$. One can prove that

$$M \subseteq \{{p_0}^{a_0} \cdots {p_n}^{a_n} \mid a_0 < l_0, \cdots, a_n < l_n\}.$$

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 $x \in M$, x is ${p_0}^{a_0} \cdots {p_n}^{a_n}$, but if there is a k such that $a_k \geqslant l_k$, $x \in {p_k}^{l_k}\mathbb{N} \subseteq L\mathbb{N}$.

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

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

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

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

That consists of at most $\left(\frac{t}{ln2} + 1\right)^{n+1}$ different (n+1)-tuples:

$$\left|M \cap [1, e^t[\right| < \left(\frac{t}{ln2} + 1\right)^{n+1} \quad \Rightarrow \quad \frac{\left|M \cap [1, e^t[\right|}{e^t} < \frac{\left(\frac{t}{ln2} + 1\right)^{n+1}}{e^t} \overset{t \to \infty}{\longrightarrow} 0$$

The density of M has the limit zero, and thanks to Proposition 24 (M is infinite), we conclude that $M \notin \mathrm{Rat}(\mathbb{N})$, then $L\mathbb{N} \notin \mathrm{Rat}(\mathbb{N})$.
Finally, by contraposition

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

■

Now we prove the correctness of the algorithm for all cases.
Proof: (Theorem 31) 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 32
3- see Lemma 33
4- Else, we obtain $Inv(m, q) \cap P = \emptyset$. Each element of P have a (strict) common divisor with q, those of $P + q\mathbb{N}$ too, and those of $(P + q\mathbb{N})\mathbb{N}$ also.
$\forall x \in P, \exists u \in \mathcal{P}, u \mid q$ and $u \mid x$, consequently $x\mathbb{N} \subseteq u\mathbb{N}$

$$\Rightarrow \; P\mathbb{N} = \bigcup_{x \in P} x\mathbb{N} \subseteq \bigcup_{u \in \mathcal{P}, u \mid q, \exists x \in P, u \mid x} u\mathbb{N}$$

Let M be the last set $\left(\bigcup_{u \in \mathcal{P}, u \mid q, \exists x \in P, u \mid x} u\mathbb{N}\right)$, we also have $(P + q\mathbb{N})\mathbb{N} \subseteq M$. The set M is rational, because it is a finite union of rational sets ($u \mid q$). Then

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

But $(P + q\mathbb{N})\mathbb{N} \subseteq M$, and we already know that $(\mathbb{N} \setminus M) \cap L\mathbb{N} = (\mathbb{N} \setminus M) \cap I\mathbb{N}$ is rational (I is finite). So we have

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

We decompose into:

$$M \cap L\mathbb{N} = \left(\bigcup_{u \in \mathcal{P}, u \mid q, \exists x \in P, u \mid x} u\mathbb{N}\right) \cap L\mathbb{N} = \bigcup_{u \in \mathcal{P}, u \mid q, \exists x \in P, u \mid x} (u\mathbb{N} \cap L\mathbb{N});$$

$$L\mathbb{N} \in \mathrm{Rat}(\mathbb{N}) \;\Leftrightarrow\; \forall u \in \mathcal{P}\Big(u \mid q \wedge \exists x \in P, u \mid x \Rightarrow (u\mathbb{N} \cap L\mathbb{N}) \in \mathrm{Rat}(\mathbb{N})\Big)$$

We still have to calculate the $u\mathbb{N} \cap L\mathbb{N}$. Let $u \in \mathcal{P}, u \mid q$, and $x \in I$.By definition of the lcm, $u\mathbb{N} \cap x\mathbb{N} = \mathrm{lcm}(u, x)\mathbb{N}$ (Appendix, Prop. 51). That is to say, if $u \mid x$, then $u\mathbb{N} \cap x\mathbb{N} = x\mathbb{N}$, and else u is relatively prime to x ($u \in \mathcal{P}$), $u\mathbb{N} \cap x\mathbb{N} = ux\mathbb{N}$. We can already conclude that

$$u\mathbb{N} \cap I\mathbb{N} = \bigcup_{x \in I} u\mathbb{N} \cap x\mathbb{N} = \bigcup_{x \in I} \mathrm{lcm}(u, x)\mathbb{N}.$$

Otherwise, for all $x \in P$, if $u \mid x$, then u divides each element of $x + q\mathbb{N}$, and $u\mathbb{N} \cap (x + q\mathbb{N})\mathbb{N} = (x + q\mathbb{N})\mathbb{N}$. Else $\gcd(u, x) = 1$ and u is also relatively prime to each integer of $x + q\mathbb{N}$, in this case $u\mathbb{N} \cap (x + q\mathbb{N})\mathbb{N} = (x + q\mathbb{N})u\mathbb{N}$. Globally

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

Following Proposition 51 of the appendix, we have $\frac{\mathrm{lcm}(u,x)}{u} = \frac{x}{\gcd(u,x)}$.
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 21, 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 all the new elements are smaller than $(x+q)/u$, and

$$\frac{x+q}{u} < \frac{m+q+q}{u} \leqslant \frac{m}{2} + q \leqslant m+q$$

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

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

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

Let $p \in \mathcal{P}$, $p \leqslant m+q$. 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: $p^{k+1}, \cdots, p^{k+q} \pmod q$ (the following values are the same 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: $\forall A, B \subseteq \mathbb{N},\ A, B \neq \emptyset$,

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

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 $u \in \mathcal{P}, u \mid q, \exists x \in P,\ u \mid x$. Let y be the smallest of the $x \in I \cup P$ such that $u \mid x$. By construction, $\forall x \in I \cup P,\ x < y \Rightarrow u \nmid x$, we find again x (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 at 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 a well-order (it is not the case in general). 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') and there is no cycle. ■

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 the power of rational languages is very often *not* rational. For example

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

(It is easier to factorize at first with $\gcd(a,b)$).

4 Conclusion

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

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

Depending on the presence of relative prime numbers in M, our algorithm either states directly that M is not rational using Dirichlet's theorem, or we consider the representation of all prime numbers in M, or we decompose M into "smaller" languages by intersection with simple periodic sets and use recursivity. It was necessary to use some facts of arithmetic, and the computation could be quite long for some cases. We do not perform here an exact computation of running time.

As future work, one could try to extend the result of Theorem 31 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: $(2+\mathbb{N})(2+\mathbb{N}) = (4+\mathbb{N}) \setminus \mathcal{P} \;\notin \mathrm{Rat}(\mathbb{N})$. 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 $M \in \mathrm{Rat}(\mathbb{N})$,

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

5 Acknowledgement

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.

[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.

Appendix: Facts of arithmetic

We recall some results of arithmetic that are necessary for our proofs. They can be found e.g. in [AF 87] or [GKP 94] (Chap.4). The main fact we have used is Theorem 52 below.

We write "$a \mid b$" for "a divides b". We assume that $a \mid 0$, even if $a = 0$, and $\gcd(0, a) = a$ if $a > 0$. The set of prime numbers is denoted by $\mathcal{P}$. It is well known that $\mathcal{P}$ is infinite and not rational. The idea of the classical proof will be reused in Lemma 32. 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 51 *For all $a, b \in \mathbb{N}$,*

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

Proof: The first formula is straightforward.
For the second one, we decompose a and b into product of prime factors. Let $(p_k)_{k \geqslant 0}$ be the sequel of the prime numbers ($\mathcal{P}$). We write

$$a = \prod_{k \geqslant 0} {p_k}^{a_k}, \quad b = \prod_{k \geqslant 0} {p_k}^{b_k}.$$

Then $\gcd(a,b) = \prod_{k \geqslant 0} {p_k}^{\min(a_k, b_k)}$, and $\mathrm{lcm}(a,b) = \prod_{k \geqslant 0} {p_k}^{\max(a_k, b_k)}$. We find again in $\mathrm{lcm}(a,b)\gcd(a,b)$ the factors p^k of a and b, redistributed. ■

Theorem 52 (Dirichlet 1840) *For all $a, b > 0$,*

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

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.

Aachener Informatik-Berichte

This is a list of recent technical reports. To obtain copies of technical reports please consult http://aib.informatik.rwth-aachen.de/ or send your request to: Informatik-Bibliothek, RWTH Aachen, Ahornstr. 55, 52056 Aachen, Email: `biblio@informatik.rwth-aachen.de`

95-11 *	M. Staudt / K. von Thadden: Subsumption Checking in Knowledge Bases
95-12 *	G.V. Zemanek / H.W. Nissen / H. Hubert / M. Jarke: Requirements Analysis from Multiple Perspectives: Experiences with Conceptual Modeling Technology
95-13 *	M. Staudt / M. Jarke: Incremental Maintenance of Externally Materialized Views
95-14 *	P. Peters / P. Szczurko / M. Jeusfeld: Business Process Oriented Information Management: Conceptual Models at Work
95-15 *	S. Rams / M. Jarke: Proceedings of the Fifth Annual Workshop on Information Technologies & Systems
95-16 *	W. Hans / St. Winkler / F. Sáenz: Distributed Execution in Functional Logic Programming
96-1 *	Jahresbericht 1995
96-2	M. Hanus / Chr. Prehofer: Higher-Order Narrowing with Definitional Trees
96-3 *	W. Scheufele / G. Moerkotte: Optimal Ordering of Selections and Joins in Acyclic Queries with Expensive Predicates
96-4	K. Pohl: PRO-ART: Enabling Requirements Pre-Traceability
96-5	K. Pohl: Requirements Engineering: An Overview
96-6 *	M. Jarke / W. Marquardt: Design and Evaluation of Computer–Aided Process Modelling Tools
96-7	O. Chitil: The ς-Semantics: A Comprehensive Semantics for Functional Programs
96-8 *	S. Sripada: On Entropy and the Limitations of the Second Law of Thermodynamics
96-9	M. Hanus (Ed.): Proceedings of the Poster Session of ALP'96 — Fifth International Conference on Algebraic and Logic Programming
96-10	R. Conradi / B. Westfechtel: Version Models for Software Configuration Management
96-11 *	C. Weise / D. Lenzkes: A Fast Decision Algorithm for Timed Refinement
96-12 *	R. Dömges / K. Pohl / M. Jarke / B. Lohmann / W. Marquardt: PRO-ART/CE* — An Environment for Managing the Evolution of Chemical Process Simulation Models
96-13 *	K. Pohl / R. Klamma / K. Weidenhaupt / R. Dömges / P. Haumer / M. Jarke: A Framework for Process-Integrated Tools

96-14 * R. Gallersdörfer / K. Klabunde / A. Stolz / M. Eßmajor: INDIA — Intelligent Networks as a Data Intensive Application, Final Project Report, June 1996

96-15 * H. Schimpe / M. Staudt: VAREX: An Environment for Validating and Refining Rule Bases

96-16 * M. Jarke / M. Gebhardt, S. Jacobs, H. Nissen: Conflict Analysis Across Heterogeneous Viewpoints: Formalization and Visualization

96-17 M. Jeusfeld / T. X. Bui: Decision Support Components on the Internet

96-18 M. Jeusfeld / M. Papazoglou: Information Brokering: Design, Search and Transformation

96-19 * P. Peters / M. Jarke: Simulating the impact of information flows in networked organizations

96-20 M. Jarke / P. Peters / M. Jeusfeld: Model-driven planning and design of cooperative information systems

96-21 * G. de Michelis / E. Dubois / M. Jarke / F. Matthes / J. Mylopoulos / K. Pohl / J. Schmidt / C. Woo / E. Yu: Cooperative information systems: a manifesto

96-22 * S. Jacobs / M. Gebhardt, S. Kethers, W. Rzasa: Filling HTML forms simultaneously: CoWeb architecture and functionality

96-23 * M. Gebhardt / S. Jacobs: Conflict Management in Design

97-01 Jahresbericht 1996

97-02 J. Faassen: Using full parallel Boltzmann Machines for Optimization

97-03 A. Winter / A. Schürr: Modules and Updatable Graph Views for PROgrammed Graph REwriting Systems

97-04 M. Mohnen / S. Tobies: Implementing Context Patterns in the Glasgow Haskell Compiler

97-05 * S. Gruner: Schemakorrespondenzaxiome unterstützen die paargrammatische Spezifikation inkrementeller Integrationswerkzeuge

97-06 M. Nicola / M. Jarke: Design and Evaluation of Wireless Health Care Information Systems in Developing Countries

97-07 P. Hofstedt: Taskparallele Skelette für irregulär strukturierte Probleme in deklarativen Sprachen

97-08 D. Blostein / A. Schürr: Computing with Graphs and Graph Rewriting

97-09 C.-A. Krapp / B. Westfechtel: Feedback Handling in Dynamic Task Nets

97-10 M. Nicola / M. Jarke: Integrating Replication and Communication in Performance Models of Distributed Databases

97-13 M. Mohnen: Optimising the Memory Management of Higher-Order Functional Programs

97-14 R. Baumann: Client/Server Distribution in a Structure-Oriented Database Management System

97-15 G. H. Botorog: High-Level Parallel Programming and the Efficient Implementation of Numerical Algorithms

98-01 * Jahresbericht 1997

98-02 S. Gruner/ M. Nagel / A. Schürr: Fine-grained and Structure-oriented Integration Tools are Needed for Product Development Processes

98-03 S. Gruner: Einige Anmerkungen zur graphgrammatischen Spezifikation von Integrationswerkzeugen nach Westfechtel, Janning, Lefering und Schürr

98-04 * O. Kubitz: Mobile Robots in Dynamic Environments

98-05 M. Leucker / St. Tobies: Truth — A Verification Platform for Distributed Systems

98-07 M. Arnold / M. Erdmann / M. Glinz / P. Haumer / R. Knoll / B. Paech / K. Pohl / J. Ryser / R. Studer / K. Weidenhaupt: Survey on the Scenario Use in Twelve Selected Industrial Projects

98-08 * H. Aust: Sprachverstehen und Dialogmodellierung in natürlichsprachlichen Informationssystemen

98-09 * Th. Lehmann: Geometrische Ausrichtung medizinischer Bilder am Beispiel intraoraler Radiographien

98-10 * M. Nicola / M. Jarke: Performance Modeling of Distributed and Replicated Databases

98-11 * A. Schleicher / B. Westfechtel / D. Jäger: Modeling Dynamic Software Processes in UML

98-12 * W. Appelt / M. Jarke: Interoperable Tools for Cooperation Support using the World Wide Web

98-13 K. Indermark: Semantik rekursiver Funktionsdefinitionen mit Striktheitsinformation

99-01 * Jahresbericht 1998

99-02 * F. Huch: Verifcation of Erlang Programs using Abstract Interpretation and Model Checking — Extended Version

99-03 * R. Gallersdörfer / M. Jarke / M. Nicola: The ADR Replication Manager

99-04 M. Alpuente / M. Hanus / S. Lucas / G. Vidal: Specialization of Functional Logic Programs Based on Needed Narrowing

99-07 Th. Wilke: CTL+ is exponentially more succinct than CTL

99-08 O. Matz: Dot-Depth and Monadic Quantifier Alternation over Pictures

2000-01 * Jahresbericht 1999

2000-02 Jens Vöge / Marcin Jurdziński: A Discrete Strategy Improvement Algorithm for Solving Parity Games

2000-04 Andreas Becks, Stefan Sklorz, Matthias Jarke: Exploring the Semantic Structure of Technical Document Collections: A Cooperative Systems Approach

2000-05 * Mareike Schoop: Cooperative Document Management

2000-06 * Mareike Schoop, Christoph Quix (Ed.): Proceedings of the Fifth International Workshop on the Language-Action Perspective on Communication Modelling

2000-07 * Markus Mohnen / Pieter Koopman (Eds.): Proceedings of the 12th International Workshop of Functional Languages

2000-08 Thomas Arts / Thomas Noll: Verifying Generic Erlang Client-Server Implementations

2001-01 * Jahresbericht 2000

2001-02 Benedikt Bollig / Martin Leucker: Deciding LTL over Mazurkiewicz Traces

2001-03 Thierry Cachat: The power of one-letter rational languages

2001-04 Benedikt Bollig / Martin Leucker / Michael Weber: Local Parallel Model Checking for the Alternation free μ-calculus

2001-05 Benedikt Bollig / Martin Leucker / Thomas Noll: Regular MSC languages

* These reports are only available as a printed version.

Please contact `biblio@informatik.rwth-aachen.de` to obtain copies.