

Infinitely generated semigroups and polynomial complexity

J.C. Birget

Abstract

This paper continues the functional approach to the P-versus-NP problem, begun in [2]. Here we focus on the monoid \mathcal{RM}_2^P of right-ideal morphisms of the free monoid, that have polynomial input balance and polynomial time-complexity. We construct a machine model for the functions in \mathcal{RM}_2^P , and evaluation functions. We prove that \mathcal{RM}_2^P is not finitely generated, and use this to show separation results for time-complexity.

1 Introduction

In [2] we defined the monoids of partial functions \mathbf{fP} and \mathcal{RM}_2^P . The question whether $P = NP$ is equivalent to the question whether these monoids are regular. The monoid \mathbf{fP} consists of all partial functions $A^* \rightarrow A^*$ that are computable by deterministic Turing machines in polynomial time, and that have polynomial I/O-balance. The submonoid \mathcal{RM}_2^P consists of the elements of \mathbf{fP} that are right-ideal morphisms of A^* . One-way functions (according to worst-case time-complexity) are exactly the non-regular elements of \mathbf{fP} . It is known that one-way functions (according to worst-case time-complexity) exist iff $P \neq NP$. Also, $f \in \mathcal{RM}_2^P$ is regular in \mathbf{fP} iff f is regular in \mathcal{RM}_2^P . Hence, $P = NP$ iff \mathbf{fP} is regular, iff \mathcal{RM}_2^P is regular. We refer to [9, 15] for background on P and NP .

The original motivation for studying \mathcal{RM}_2^P in addition to \mathbf{fP} was that \mathcal{RM}_2^P is reminiscent of the Thompson-Higman groups [14, 16, 12, 7, 6, 5] and the Thompson-Higman monoids [4]. It also quickly turned out that \mathcal{RM}_2^P , while having the same connection to P-vs.-NP as \mathbf{fP} , has different properties than \mathbf{fP} (e.g., regarding the Green relations, and actions on $\{0, 1\}^\omega$; see [2, 3]). It is hard to know whether this approach will contribute to a solution of the P-vs.-NP problem, but the monoids \mathbf{fP} and \mathcal{RM}_2^P are interesting by themselves.

Above and in the rest of the paper we use the following notation and terminology. We have an alphabet A , which will be $\{0, 1\}$ unless the contrary is explicitly stated, and A^* denotes the set of all strings over A , including the empty string ε . For $x \in A^*$, $|x|$ denotes the length of the string x . For a partial function $f : A^* \rightarrow A^*$, the domain is $\text{Dom}(f) = \{x \in A^* : f(x) \text{ is defined}\}$, and the image is $\text{Im}(f) = f(A^*) = f(\text{Dom}(f))$. When we say “function”, we mean partial function (except when we explicitly say “total function”). Similarly, for a deterministic input-output Turing machine with input-output alphabet A , the domain of the machine is the set of input words for which the machine produces an output; and the set of output words is the image of the machine.

A function $f : A^* \rightarrow A^*$ is called *polynomially balanced* iff there exists polynomials p, q such that for all $x \in \text{Dom}(f)$: $|f(x)| \leq p(|x|)$ and $|x| \leq q(|f(x)|)$. The polynomial q is called an *input balance* function for f .

As we said already, \mathbf{fP} is the set of partial functions $f : A^* \rightarrow A^*$ that are polynomially balanced, and such that $x \in \text{Dom}(f) \mapsto f(x)$ is computable by a deterministic polynomial-time Turing machine. Hence, $\text{Dom}(f)$ is in P when $f \in \mathbf{fP}$, and it is not hard to show that $\text{Im}(f)$ is in NP . Clearly, \mathbf{fP} is a monoid under function composition.

A function $f : A^* \rightarrow A^*$ is said to be *one-way (with respect to worst-case complexity)* iff $f \in \mathbf{fP}$, but there exists *no* deterministic polynomial-time algorithm which, on every input $y \in \text{Im}(f)$, outputs some $x \in A^*$ such that $f(x) = y$. By “one-way” we will always mean one-way with respect to worst-case complexity; hence, these functions are not “cryptographic one-way functions” (in the sense of, e.g.,

[8, 13, 10]). However, they are important for the P-vs.-NP problem because of the following folklore fact (see e.g., [11] p. 33): *One-way functions exist iff P ≠ NP*.

As is easy to prove (see the Introduction of [2]), $f \in \mathbf{fP}$ is not one-way iff f is *regular* in \mathbf{fP} . By definition, an element f in a monoid M is regular iff there exists $f' \in M$ such that $ff'f = f$; in this case, f' is called an *inverse* of f .¹ A monoid M is called regular iff all the elements of M are regular. In summary we have: *The monoid \mathbf{fP} is regular iff P = NP*.

Let us look in more detail at the monoid \mathcal{RM}_2^P . A *right ideal* of A^* is a subset $R \subseteq A^*$ such that $RA^* = R$ (i.e., R is closed under right-concatenation by any string). For two strings $v, w \in A^*$, we say that v is a *prefix* of w iff $(\exists x \in A^*)[vx = w]$. A *prefix code* in A^* is a set $P \subset A^*$ such that no word in P is a prefix of another word in P . For any right ideal R there exists a unique prefix code P_R such that $R = P_R A^*$; we say that P_R generates R as a right ideal. For details, see e.g. [6, 4]; a good reference on prefix codes, and variable-length codes in general is [1].

A *right-ideal morphism* is a partial function $h : A^* \rightarrow A^*$ such that for all $x \in \text{Dom}(h)$ and all $w \in A^*$: $h(xw) = h(x)w$. In that case, $\text{Dom}(h)$ and $\text{Im}(h)$ are right ideals. For a right-ideal morphism h , let $\text{domC}(h)$ (called the *domain code*) be the prefix code that generates $\text{Dom}(h)$ as a right ideal. Similarly, let $\text{imC}(h)$, called the *image code*, be the prefix code that generates $\text{Im}(h)$. So a right-ideal morphism h is determined by $h|_{\text{domC}(h)}$ (the restriction of h to its domain code). In general, $\text{imC}(h) \subseteq h(\text{domC}(h))$, and it can happen that $\text{imC}(h) \neq h(\text{domC}(h))$. We define

$$\mathcal{RM}_2^P = \{f \in \mathbf{fP} : f \text{ is a right-ideal morphism of } A^*\}.$$

By Prop. 2.6 in [2], $f \in \mathcal{RM}_2^P$ is regular in \mathcal{RM}_2^P iff f is regular in \mathbf{fP} . *The monoid \mathcal{RM}_2^P is regular iff P = NP*.

We saw (Cor. 2.9 in [2]) that \mathbf{fP} and \mathcal{RM}_2^P are not isomorphic, that the group of units of \mathcal{RM}_2^P is trivial (Prop. 2.12 in [2]), and that \mathcal{RM}_2^P has only one non-0 \mathcal{J} -class (Prop. 2.7 in [2]). In [3] we will see that \mathcal{RM}_2^P has interesting actions on $\{0, 1\}^\omega$, and has interesting homomorphic images (some of which are regular monoids, and some of which are regular iff P = NP). Overall, \mathcal{RM}_2^P seems to have “more structure” than \mathbf{fP} .

It is proved in [2] (Section 3) that \mathbf{fP} is isomorphic to a submonoid of \mathcal{RM}_2^P . To prove this, we use an encoding of the three-letter alphabet $\{0, 1, \#\}$ into words over the two-letter alphabet $\{0, 1\}$; this encoding will also be used here. First, we encode the alphabet $\{0, 1, \#\}$ by $\text{code}(0) = 00$, $\text{code}(1) = 01$, $\text{code}(\#) = 11$. A word $x_1 \dots x_n \in \{0, 1, \#\}^*$ is encoded to $\text{code}(x_1) \dots \text{code}(x_n)$. For a fixed $k > 0$, a k -tuple of words $(u_1, \dots, u_{k-1}, u_k) \in \{0, 1\}^* \times \dots \times \{0, 1\}^*$ is encoded to $\text{code}(u_1 \# \dots u_{k-1} \#) u_k = \text{code}(u_1) 11 \dots \text{code}(u_{k-1}) 11 u_k \in \{0, 1\}^*$. A function $f \in \mathbf{fP}$ is encoded to $f^C \in \mathcal{RM}_2^P$, defined by $\text{domC}(f^C) = \text{code}(\text{Dom}(f) \#)$, so $\text{Dom}(f^C) = \text{code}(\text{Dom}(f)) 11 \{0, 1\}^*$; and

$$f^C(\text{code}(x \#) v) = \text{code}(f(x) \#) v,$$

for all $x \in \text{Dom}(f)$ and $v \in \{0, 1\}^*$; equivalently, $f^C(\text{code}(x) 11 v) = \text{code}(f(x)) 11 v$. Then for every $L \subseteq \{0, 1\}^*$, $\text{code}(L \#)$ is a prefix code, which belongs to P iff L is in P. And $f \in \mathbf{fP}$ iff $f^C \in \mathcal{RM}_2^P$. The transformation $f \mapsto f^C$ is a isomorphic embedding of \mathbf{fP} into \mathcal{RM}_2^P ; moreover, f^C is regular in \mathcal{RM}_2^P iff f is regular in \mathbf{fP} . From here on, the alphabet denoted by A will always be $\{0, 1\}$.

In [2] (Section 4) we introduced a notion of *polynomial program* for Turing machines with built-in polynomial counter (for input balance and time-complexity). These programs form a machine model that characterizes the functions in \mathbf{fP} . For a polynomial program w , we let $\phi_w \in \mathbf{fP}$ denote the function computed by this program. For every polynomial q of the form $q(n) = an^k + a$ (where a, k are positive integers), we constructed an *evaluation map* $\text{ev}_q^C \in \mathbf{fP}$ such that for every polynomial program w with built-in polynomial $p_w(n) \leq q(n)$ (for all $n \geq 0$), and all $x \in A^*$,

$$\text{ev}_q^C(\text{code}(w) 11 x) = \text{code}(w) 11 \phi_w(x)$$

¹The terminology varies, depending on the field. In semigroup theory f' such that $ff'f = f$ is called a semi-inverse or a pseudo-inverse of f , in numerical mathematics f' is called a generalized inverse, in ring theory and in category theory it's called a weak inverse. In semigroup theory the term “inverse” of f is only applied to f' if $f'ff' = f'$ holds in addition to $ff'f = f$. It is easy to see that if $ff'f = f$ then $f'ff' (= v)$ satisfies $fvf = f$ and $vvf = v$.

if $x \in \text{Dom}(\phi_w)$; if $x \notin \text{Dom}(\phi_w)$ then $\text{ev}_q^C(\text{code}(w)11x)$ is undefined. We used ev_q^C , with any polynomial q of degree ≥ 2 with large enough coefficient, to prove the following: First, fP is finitely generated (Theorem 4.5 in [2]). Second, ev_q^C is complete in fP with respect to inversive polynomial reduction (Section 5 of [2]). Later in this paper (Def. 2.4 and following) we define completeness and various reductions for \mathcal{RM}_2^P , along the same lines as for fP .

Note that fP and \mathcal{RM}_2^P , in their entirety, do not have evaluation maps that belong to fP , respectively \mathcal{RM}_2^P (since such maps would not have polynomially bounded complexity). That is the reason why we restrict ev and evR to complexity $\leq q(\cdot)$, and why we need precise machine models for fP and \mathcal{RM}_2^P (as opposed to more intuitive “higher-level” models).

In Section 2 we define a machine model that characterizes the functions in \mathcal{RM}_2^P ; and for any large enough polynomial q we construct evaluation maps evR_q^C and evR_q^{CC} for the functions in \mathcal{RM}_2^P that have balance and time-complexity $\leq q$. We prove that evR_q^{CC} is complete in \mathcal{RM}_2^P (and in fP) with respect to inversive Turing reduction. In Section 3 we prove that \mathcal{RM}_2^P is not finitely generated, and in Section 4 we show that infinite generation has some complexity consequences, i.e., infinite generation can be used for a time-complexity lower-bound argument.

2 Machine model and evaluation maps for \mathcal{RM}_2^P

The evaluation map $\text{ev}_q^C : \text{code}(w)11x \mapsto \text{code}(w)11\phi_w(x)$, that we constructed for fP in [2], works in particular when $\phi_w \in \mathcal{RM}_2^P$ (provided that ϕ_w has time-complexity and input-balance $\leq q$). But ev_q^C is not a right-ideal morphism and, moreover, ev_q^C can evaluate functions that are not in \mathcal{RM}_2^P . We want to construct an evaluation map that belongs to \mathcal{RM}_2^P , and that evaluates exactly the elements of \mathcal{RM}_2^P that have balance and complexity $\leq q$. In [2] we constructed a machine model for fP , namely a class of Turing machines with built-in polynomial counter (for controlling the time-complexity and the input-balance). We will refine these Turing machines in order to obtain a machine model for accepting the right ideals in P , and for computing the functions in \mathcal{RM}_2^P .

We will consider deterministic multi-tape Turing machines with input-output alphabet A , with a read-only input tape, and a write-only output tape. Moreover we assume that on the input tape and on the output tape, the head can only move to the right, or stay in place (but cannot move left). We assume that the input tape has a left endmarker $\#$, and a right endmarker B (the blank symbol). At the beginning of a computation of such a machine M on input $z \in A^*$, the input tape has content $\#zB$, with the input tape head on $\#$; initially, all other tapes are blank (i.e., they are filled with infinitely many copies of the letter B). The output tape does not need endmarkers (since it is write-only). We assume that M has a special *output state* q_{out} , and that M only goes to state q_{out} when the output is complete; the output state is a halting state (i.e., M has no transition from state q_{out}). An important convention for a Turing machine M with non-total input-output function f_M is the following: *If M on input x halts in a state that is not q_{out} , then there is no output (even if the output tape contains a non-blank word)*. So, in that case, $f_M(x)$ is undefined. The content of the output tape is considered unreadable, or hidden, until the output state q_{out} is reached.

This kind of Turing machine can compute any partial recursive function (the restrictions on the input and output tapes do not limit the machine, because of the work-tapes). To compute a function in fP , we add a built-in polynomial (used as a bound on input balance and time-complexity); see Section 4 in [2].

In order to obtain a machine model for the functions in \mathcal{RM}_2^P the above Turing machines (with built-in polynomial) will be restricted so that they compute right-ideal morphisms of A^* . This is done in two steps: First, sequential functions and sequential Turing machines are introduced. From this it is easy to obtain a class of Turing machines that compute right-ideal morphisms (which are a special kind of sequential functions). Recall that by “function” we mean partial function. By definition, a function $f : A^* \rightarrow A^*$ is *sequential* iff

for all $x_1, x_2 \in \text{Dom}(f)$: if x_1 is a prefix of x_2 then $f(x_1)$ is a prefix of $f(x_2)$.

Obviously, every right-ideal morphism is a sequential function.

A *sequential Turing machine* is a deterministic multi-tape Turing machine M (with special input tape and special output tape and output state, according to the conventions above), with input-output function f_M , such that the following holds.

For every $x \in \text{Dom}(f_M)$ and every word $z \in A^$: in the computation of M on input xz , the input-tape head does not start reading zB until $f_M(x)$ has been written on the output tape.*

To “read a letter ℓ ” (in zB) means to make a transition whose input letter is ℓ . So, the input tape has content $\#xzB$, with the input-tape head on the left-most letter of zB (but no transition has been made on that letter yet), and the output tape now has content $f_M(x)$. Of course, at this moment the computation of M on input xz is not necessarily finished; the state is not necessarily q_{out} , the output might still grow, and q_{out} might be reached eventually, or not; if q_{out} is never reached, there is no final output.

The sequential Turing machines form a machine model for the partial recursive sequential functions. If we let the machines have a built-in polynomial we obtain a machine model for the sequential functions in fP .

Finally, to obtain a machine model for the functions in \mathcal{RM}_2^P we take the sequential Turing machines with built-in polynomial, with the following additional condition.

For every $x \in \text{Dom}(f_M)$ and every word $z \in A^$: in the computation of M on input xz , once $f_M(x)$ has been written on the output tape (after x was read on the input tape), the remaining input z is copied to the output tape; at this point the state q_{out} is reached.*

We call such a machine an \mathcal{RM}_2^P -machine.

The following shows how, from an fP -machine for a function f , an \mathcal{RM}_2^P -machine for f can be constructed, provided that $f \in \mathcal{RM}_2^P$.

Let us first consider right ideals in P , rather than functions. For any polynomial program w for a Turing machine M_w that accepts a language $L \in P$, we construct a new polynomial program v describing a Turing machine M_v that behaves as follows: On any input $x \in \{0,1\}^*$, M_v successively examines prefixes of x until it finds a prefix, say p , that is accepted by M_w ; M_v does not read the letter of x that comes after p until it has decided that $p \notin L$. As soon as M_v finds a prefix p of x such that $p \in L$, M_v accepts the whole input x . If M_w accepts no prefix of x , M_v rejects x . Thus, M_v accepts LA^* (the right ideal generated by L); if L is a right ideal then $LA^* = L$. If M_w has time-complexity $\leq T(\cdot)$ (a polynomial) then M_v has time-complexity $\leq T(\cdot)^2$.

Let us now consider functions in \mathcal{RM}_2^P . Given any polynomial program w for a function $\phi_w \in \text{fP}$, we construct a new polynomial program v such that M_v , on input x , successively examines all prefixes of x until it finds a prefix p in $\text{Dom}(\phi_w)$; let $\phi_w(p) = y$. Then, on input x , the machine M_v outputs yz , where z is such that $x = pz$. Note that since p is the shortest prefix of x such that $p \in \text{Dom}(\phi_w)$, we actually have $p \in \text{domC}(\phi_w)$ (if $\text{Dom}(\phi_w)$ is a right ideal). The machine M_v does not read the letter of x that comes after a prefix p until it has decided that $p \notin \text{Dom}(\phi_w)$ or $p \in \text{domC}(\phi_w)$. Hence, the function computed by M_v is in \mathcal{RM}_2^P . This construction describes a transformation $f \in \text{fP} \mapsto f_{\text{pref}} \in \mathcal{RM}_2^P$, where f_{pref} is defined as follows:

$$f_{\text{pref}}(x) = f(p)z,$$

where $x = pz$, and p is the shortest prefix of x that belongs to $\text{Dom}(f)$; so, $p \in \text{domC}(f_{\text{pref}})$. Thus for every $f \in \text{fP}$ we have: $f \in \mathcal{RM}_2^P$ iff $f_{\text{pref}} = f$.

Based on \mathcal{RM}_2^P -machines we can construct evaluation maps for \mathcal{RM}_2^P . Let q be a polynomial where $q(n) = an^k + a$ for some integers $a, k \geq 1$. We define evR_q^C , as follows:

$$\text{evR}_q^C(\text{code}(w) \ 11 \ x) = \text{code}(w) \ 11 \ \phi_w(x),$$

for all \mathcal{RM}_2^P -programs w with built-in polynomial $p_w \leq q$, and for all $x \in \text{Dom}(\phi_w)$. The details of the construction are the same as for ev_q^C ; see Section 4 in [2]. Although evR_q^C belongs to \mathcal{RM}_2^P and

evaluates all \mathcal{RM}_2^P -programs w with built-in polynomial $\leq q$, we will prove in Theorem 4.3 that the complexity of evR_q^C is higher than q .

The following *doubly coded evaluation function* is usually more useful for \mathcal{RM}_2^P -programs. It is defined by

$$\text{evR}_q^{CC}(\text{code}(w) \text{ } 11 \text{ } \text{code}(u) \text{ } 11 \text{ } v) = \text{code}(w) \text{ } 11 \text{ } \text{code}(\phi_w(u)) \text{ } 11 \text{ } v,$$

when $u \in \text{domC}(\phi_w)$, $v \in A^*$, and w is as before.

To give a relation between evR_q^C and evR_q^{CC} we will use the following partial recursive right-ideal morphism γ , defined for very \mathcal{RM}_2^P -program $w \in A^*$ and every $x \in \text{Dom}(\phi_w)$ by

$$\gamma(\text{code}(w) \text{ } 11 \text{ } x) = \text{code}(w) \text{ } 11 \text{ } \text{code}(u) \text{ } 11 \text{ } v,$$

where $x = uv$, and u is the shortest prefix of x such that $u \in \text{Dom}(\phi_w)$; equivalently, $u \in \text{domC}(\phi_w)$. When $x \notin \text{Dom}(\phi_w)$, $\gamma(\text{code}(w) \text{ } 11 \text{ } x)$ is undefined. Essentially, γ finds the shortest prefix of x that belongs to $\text{Dom}(\phi_w)$ (or equivalently, to $\text{domC}(\phi_w)$). The function γ can be evaluated by examining successively longer prefixes of x until a prefix $u \in \text{Dom}(\phi_w)$ is found. So γ is computable with recursive domain, when w ranges over \mathcal{RM}_2^P -programs.

For any fixed \mathcal{RM}_2^P -program w , let γ_w be γ restricted to this w , i.e., $\gamma_w = \gamma|_{\text{code}(w) \text{ } 11 \text{ } A^*}$. In other words, $\text{Dom}(\gamma_w) = \text{code}(w) \text{ } 11 \text{ } \text{Dom}(\phi_w)$, and

$$\gamma_w(\text{code}(w) \text{ } 11 \text{ } uv) = \text{code}(w) \text{ } 11 \text{ } \text{code}(u) \text{ } 11 \text{ } v$$

when $u \in \text{domC}(\phi_w)$, $v \in A^*$. Similarly we define γ_w^o by $\text{Dom}(\gamma_w^o) = \text{Dom}(\phi_w)$ (as opposed to $\text{code}(w) \text{ } 11 \text{ } \text{Dom}(\phi_w)$), and

$$\gamma_w^o(uv) = \text{code}(w) \text{ } 11 \text{ } \text{code}(u) \text{ } 11 \text{ } v$$

when $u \in \text{domC}(\phi_w)$, $v \in A^*$. So, $\text{Im}(\gamma_w^o) = \text{Im}(\gamma_w) = \text{code}(w) \text{ } 11 \text{ } \text{domC}(\phi_w) \text{ } 11 \text{ } A^*$.

Then γ_w and γ_w^o belong to \mathcal{RM}_2^P for every fixed w . But γ itself is not polynomial-time computable, since it has to work for all possible \mathcal{RM}_2^P -programs w .

Another restricted form of γ that belongs to \mathcal{RM}_2^P is obtained by choosing a fixed polynomial q , and defining γ_q as the restriction of γ to the set

$$\{\text{code}(w) \text{ } 11 \text{ } x : w \text{ is a } \mathcal{RM}_2^P\text{-program with built-in polynomial } \leq q, \text{ and } x \in \text{Dom}(\phi_w)\}.$$

Hence, $\gamma_q \in \mathcal{RM}_2^P$.

We also define the functions $\pi_0, \pi_1, \rho_0, \rho_1 \in \mathcal{RM}_2^P$ by $\pi_a(x) = ax$, $\rho_a(ax) = x$, for all $x \in \{0, 1\}^*$ and $a \in \{0, 1\}$. For a word $w = a_n \dots a_1$ with $a_i \in \{0, 1\}$ we denote $\pi_{a_n} \circ \dots \circ \pi_{a_1}$ by π_w , and $\rho_{a_n} \circ \dots \circ \rho_{a_1}$ by ρ_w .

Then we have: $\gamma_w^o = \gamma_w \circ \pi_{\text{code}(w) \text{ } 11}$, and $\gamma_w = \gamma_w^o \circ \rho_{\text{code}(w) \text{ } 11}$.

Another important function in \mathcal{RM}_2^P is the decoding function, defined for any $u, v \in A^*$ by

$$\text{decode}(\text{code}(u) \text{ } 11 \text{ } v) = uv,$$

so $\text{domC}(\text{decode}) = \{00, 01\}^* \text{ } 11$, and $\text{imC}(\text{decode}) = \{\varepsilon\}$. We also define a second-coordinate decoding function, for all $u_1, u_2, v \in A^*$, by

$$\text{decode}_2(\text{code}(u_1) \text{ } 11 \text{ } \text{code}(u_2) \text{ } 11 \text{ } v) = \text{code}(u_1) \text{ } 11 \text{ } u_2 \text{ } v.$$

So, $\text{decode}_2 \in \mathcal{RM}_2^P$, $\text{domC}(\text{decode}_2) = \{00, 01\}^* \text{ } 11 \{00, 01\}^* \text{ } 11$, and $\text{imC}(\text{decode}_2) = \{00, 01\}^* \text{ } 11$.

Now we can formulate a relation between evR_q^C and evR_q^{CC} :

$$\text{evR}_q^C = \text{decode}_2 \circ \text{evR}_q^{CC} \circ \gamma_q.$$

In order to show that evR_q^{CC} is complete with respect to inversive reduction in \mathcal{RM}_2^P , we will adapt the padding and unpadding functions (defined for fP in [2], Section 4) to \mathcal{RM}_2^P . Although for

\mathcal{RM}_2^P we keep the same names as for the corresponding (un)padding functions in \mathbf{fP} , the functions are slightly different. The padding procedure begins with the function $\text{expand}(\cdot)$, defined by

$$\begin{aligned} \text{expand}(\text{code}(w) \ 11 \ \text{code}(u) \ 11 \ v) \\ = \text{code}(\text{ex}(w)) \ 11 \ 0^4|\text{code}(u)|^2+8|\text{code}(u)|+2 \ 01 \ \text{code}(u) \ 11 \ v, \end{aligned}$$

for all $u \in \text{domC}(\phi_w)$, $v \in A^*$, and \mathcal{RM}_2^P -programs w . The word $0^4|\text{code}(u)|^2+8|\text{code}(u)|+2 \ 01$ is of the form $\text{code}(s)$ for a word $s \in 0^*1$; and $0^4|\text{code}(u)|^2+8|\text{code}(u)|+2 \ 01 \ \text{code}(u)$ is also a code word, namely $\text{code}(su)$. Since 0^*1 and its subset $(00)^*01$ are prefix codes, $\text{code}(s) = 0^4|\text{code}(u)|^2+8|\text{code}(u)|+2 \ 01$ is uniquely determined as a prefix of $\text{code}(su)$.

Here, $\text{ex}(w)$ is an \mathcal{RM}_2^P -program obtained from w so that

$$\phi_{\text{ex}(w)}((00)^h \ 01 \ \text{code}(u) \ 11 \ v) = (00)^h \ 01 \ \text{code}(\phi_w(u)) \ 11 \ v,$$

for all $u \in \text{domC}(\phi_w)$, $v \in A^*$, and $h > 0$. Moreover, if $n \mapsto a n^k + a$ is the built-in polynomial of the program w then the built-in polynomial of $\text{ex}(w)$ is

$$p_e(n) = a_e n^{\lceil k/2 \rceil} + a_e, \text{ with } a_e = \max\{12, \lceil a/2^k \rceil + 1\}.$$

The detailed justification of the numbers used in the definition of expand and ex (as well as reexpand , recontr , and contr below) is given in [2], Section 4.

It is important that expand uses the prefix u of x for padding (in the format $\text{code}(u) \ 11$, where $u \in \text{domC}(\phi_w)$). If the whole input x were used for computing the amount of padding, expand would not be a right-ideal morphism. This is the reason why we introduce γ_w or γ_w^o , in order to isolate the prefix $u \in \text{domC}(\phi_w)$ of x .

We iterate expansion (padding) by applying the following function, where $\text{ex}(\cdot)$ is as above:

$$\begin{aligned} \text{reexpand}(\text{code}(\text{ex}(z)) \ 11 \ 0^k \ 01 \ \text{code}(u) \ 11 \ v) \\ = \text{code}(\text{ex}(z)) \ 11 \ 0^{4k^2+8k+2} \ 01 \ \text{code}(u) \ 11 \ v, \end{aligned}$$

where $k > 0$, $u, v \in A^*$, and z is any \mathcal{RM}_2^P -program; k is even in the context where reexpand will be used.

Repeated contraction (unpadding) is carried out by applying the following function, for $k > 0$:

$$\begin{aligned} \text{recontr}(\text{code}(\text{ex}(z)) \ 11 \ (00)^k \ 01 \ \text{code}(y) \ 11 \ v) \\ = \text{code}(\text{ex}(z)) \ 11 \ (00)^{\max\{1, \lfloor \sqrt{k}/2 \rfloor - 1\}} \ 01 \ \text{code}(y) \ 11 \ v; \end{aligned}$$

note that $\max\{1, \lfloor \sqrt{k}/2 \rfloor - 1\} \geq 1$.

The unpadding procedure ends with the application of the function

$$\text{contr}(\text{code}(\text{ex}(z)) \ 11 \ (00)^k \ 01 \ \text{code}(y) \ 11 \ v) = \text{code}(z) \ 11 \ \text{code}(y) \ 11 \ v,$$

if $2 \leq |(00)^k| = 2k \leq 4|\text{code}(y)|^2 + 8|\text{code}(y)| + 2$.

The functions $\text{expand}(\cdot)$, $\text{ex}(\cdot)$, $\text{reexpand}(\cdot)$, $\text{recontr}(\cdot)$, and $\text{contr}(\cdot)$, are undefined in the cases where no output has been specified above.

Lemma 2.1 *Let q_2 be the polynomial defined by $q_2(n) = 12n^2 + 12$. For any $\phi_w \in \mathcal{RM}_2^P$, where w is a \mathcal{RM}_2^P -program with built-in polynomial q (of the form $q(n) = a n^k + a$ for positive integers a, k), we have for all $u \in \text{domC}(\phi_w)$, $v \in A^*$:*

$$\begin{aligned} (\star) \quad \phi_w(uv) \\ = \text{decode} \circ \rho_{\text{code}(w) \ 11} \circ \text{contr} \circ \text{recontr}^{2m} \circ \text{evR}_{q_2}^{CC} \circ \text{reexpand}^m \circ \text{expand} \circ \gamma_w^o(uv) \\ = \rho_{\text{code}(w) \ 11} \circ \text{decode}_2 \circ \text{contr} \circ \text{recontr}^{2m} \circ \text{evR}_{q_2}^{CC} \circ \text{reexpand}^m \circ \text{expand} \circ \gamma_w^o(uv), \end{aligned}$$

where $m = \lceil \log_2(a + k) \rceil$.

Proof. This is similar to the proof of Prop. 4.5 in [2], with a few modifications. For $u \in \text{domC}(\phi_w)$, $v \in A^*$,

$$\begin{aligned} uv &\xrightarrow{\gamma_w^o} \text{code}(w) \text{ 11 } \text{code}(u) \text{ 11 } v \\ \xrightarrow{\text{expand}} &\text{code}(\text{ex}(w)) \text{ 11 } 0^4 |\text{code}(u)|^2 + 8 |\text{code}(u)| + 2 \text{ 01 } \text{code}(u) \text{ 11 } v \\ \xrightarrow{\text{reexpand}^m} &\text{code}(\text{ex}(w)) \text{ 11 } 0^{N_{m+1}} \text{ 01 } \text{code}(u) \text{ 11 } v, \end{aligned}$$

where $N_1 = 4 |\text{code}(u)|^2 + 8 |\text{code}(u)| + 2$, so $|0^{N_1} \text{ 01}| = (2(|\text{code}(u)| + 1))^2$; by induction, $N_i = 4 N_{i-1}^2 + 8 N_{i-1} + 2$ for $1 < i \leq 2m + 1$, and $|0^{N_i} \text{ 01}| = (2(N_{i-1} + 1))^2$. The above string, which will now be the argument of $\text{evR}_{q_2}^{CC}$, has length $> N_{m+1} + 2 + |\text{code}(u)|$, which is much larger than the time it takes to simulate the machine with program w on input u . So $\text{evR}_{q_2}^{CC}$ can now be applied correctly. Continuing the calculation,

$$\begin{aligned} \xrightarrow{\text{evR}_{q_2}^{CC}} &\text{code}(\text{ex}(w)) \text{ 11 } 0^{N_{m+1}} \text{ 01 } \text{code}(\phi_w(u)) \text{ 11 } v \\ \xrightarrow{\text{recontr}^{2m}} &\text{code}(w) \text{ 11 } 00 \text{ 01 } \text{code}(\phi_w(u)) \text{ 11 } v. \end{aligned}$$

We use $2m$ in recontr^{2m} because $\phi_w(u)$ could be much shorter than u ; but because of polynomial input balance, $|u| \leq p_w(|\phi_w(u)|)$. Note that doing more input padding than necessary does not do any harm; and recontracting (unpadding) more than needed has no effect (by the definition of recontr). Hence contr can now be applied correctly. We complete the calculation:

$$\xrightarrow{\text{contr}} \text{code}(w) \text{ 11 } \text{code}(\phi_w(u)) \text{ 11 } v \xrightarrow{\text{decode}_2} \text{code}(w) \text{ 11 } \phi_w(u) \text{ 11 } v \xrightarrow{\rho_{\text{code}(w)} \text{ 11}} \phi_w(u) \text{ 11 } v. \quad \square$$

Lemma 2.2. \mathcal{RM}_2^P has the following infinite generating set:

$$\begin{aligned} &\{\text{decode}, \rho_0, \rho_1, \pi_0, \pi_1, \text{contr}, \text{recontr}, \text{evR}_{q_2}^{CC}, \text{reexpand}, \text{expand}\} \\ &\cup \{\gamma_w : w \text{ is any } \mathcal{RM}_2^P\text{-program}\}. \end{aligned}$$

Here, decode can be replaced by decode_2 . Yet another infinite generating set of \mathcal{RM}_2^P is

$$\{\rho_0, \rho_1, \pi_0, \pi_1\} \cup \{\text{evR}_q^C : q \text{ is any polynomial of the form } q(n) = an^k + a \text{ with } a, k \in \mathbb{N}_{\geq 1}\}.$$

Proof. The first infinite generating set follows from Lemma 2.1. Recall that $\gamma_w^o = \gamma_w \circ \pi_{\text{code}(w)} \text{ 11}$. The second generating set follows in a straightforward way from the proof of Prop. 4.5 in [2]. \square

Proposition 2.3 \mathcal{RM}_2^P is generated by a set of regular elements of \mathcal{RM}_2^P .

Proof. The generators $\rho_0, \rho_1, \pi_0, \pi_1$ are easily seen to be regular. Thus, using the second infinite generating set in Lemma 2.2, it is enough to factor evR_q^C into regular elements. We have:

$$\text{evR}_q^C = \rho_{2,q} \circ E_q,$$

where E_q and $\rho_{2,q}$ are defined as follows: For every \mathcal{RM}_2^P -program w with built-in polynomial $\leq q$, and every $u \in \text{domC}(\phi_w)$ and $v \in A^*$,

$$E_q(\text{code}(w) \text{ 11 } \text{code}(u) \text{ 11 } v) = \text{code}(w) \text{ 11 } \text{code}(u) \text{ 11 } \text{code}(\phi_w(u)) \text{ 11 } v;$$

and for all $z, y, x, v \in A^*$ such that $|x| \leq q(|y|)$,

$$\rho_{2,q}(\text{code}(z) \text{ 11 } \text{code}(x) \text{ 11 } \text{code}(y) \text{ 11 } v) = \text{code}(z) \text{ 11 } \text{code}(y) \text{ 11 } v.$$

The functions are undefined otherwise. It is easy to see that E_q and $\rho_{2,q}$ have polynomial-time inversion algorithms (i.e., they are regular), and belong to \mathcal{RM}_2^P . \square

We will show now that $\text{evR}_{q_2}^{CC}$ is complete in \mathcal{RM}_2^P and in fP , with respect to a certain “inversive reduction”. We need to recall some definitions from [2] concerning reductions between functions in fP or \mathcal{RM}_2^P , and in particular, reductions that “preserve one-wayness” (inversive reductions).

Definition 2.4 Let $f_1, f_2 : A^* \rightarrow A^*$ be two polynomially balanced right-ideal morphisms.

- (1) We say that f_2 simulates f_1 (denoted by $f_1 \preceq f_2$) iff there exist $\alpha, \beta \in \mathcal{RM}_2^P$ such that $f_1 = \beta \circ f_2 \circ \alpha$.
- (2) We have a polynomial-time Turing simulation of f_1 by f_2 (denoted by $f_1 \preceq_T f_2$) iff f_1 can be computed by an oracle \mathcal{RM}_2^P -machine that can make oracle calls to f_2 ; such oracle calls can, in particular, be calls on the membership problem of $\text{Dom}(f_2)$.

In the above definition, f_1, f_2 need not be polynomial-time computable.

Since \mathcal{RM}_2^P is \mathcal{J}^0 -simple (Prop. 2.7 in [2]), every $f_1 \in \mathcal{RM}_2^P$ is simulated by every $f_2 \in \mathcal{RM}_2^P - \{0\}$ (for each of the above simulations).

Definition 2.5 (Inversive reduction). If \preceq_X is a simulation between right-ideal morphisms (e.g., as in the previous definition) then the corresponding inversive reduction is defined as follows. We say that f_1 inversively X -reduces to f_2 (denoted by $f_1 \leq_{\text{inv}, X} f_2$) iff

- (1) $f_1 \preceq_X f_2$, and
- (2) for every inverse f'_2 of f_2 there exists an inverse f'_1 of f_1 such that $f'_1 \preceq_X f'_2$; here, f'_2 and f'_1 range over all polynomially balanced right-ideal morphisms $A^* \rightarrow A^*$.

Note that \mathcal{J}^0 -simplicity (Prop. 2.7 in [2]) does not apply for inversive reduction since f'_2, f'_1 do not range over just \mathcal{RM}_2^P . One easily proves the following about polynomially balanced right-ideal morphisms f_1, f_2 (see [2], Section 5):

If $f_1 \leq_{\text{inv}, T} f_2$ and $f_2 \in \mathcal{RM}_2^P$, then $f_1 \in \mathcal{RM}_2^P$; if, in addition, f_2 is regular, then f_1 is regular (equivalently, if, in addition, f_1 is one-way, then f_2 is one-way).

Definition 2.6 A polynomially balanced right-ideal morphism f_0 is complete in a set S (of right-ideal morphisms) with respect to an (inversive) reduction $\leq_{\text{inv}, X}$ iff $f_0 \in S$, and for all $\phi \in S$: $\phi \leq_{\text{inv}, X} f_0$.

See Section 5 of [2] for more details and properties of these simulations and reductions; in [2] the focus was on fP , whereas here we concentrate on \mathcal{RM}_2^P . The simulations in Def. 2.4 are similar to the standard notions of reductions between decision problems. The concept of inversive reduction was first introduced in [2]; it is the appropriate notion of reduction between functions when one-wayness is to be preserved under upward reduction (and regularity is to be preserved under downward reduction).

In the above definitions we only refer to polynomially balanced inverses; this is justified by the following Proposition, according to which “balanced functions have balanced inverses”.

Proposition 2.7 Suppose f is a right-ideal morphism with balance $\leq q(\cdot)$ (where $q(\cdot)$ is a polynomial), and f has an inverse f'_1 with time-complexity $\leq T(\cdot)$. Then f has an inverse f' with balance $\leq q$ and time-complexity $\leq T(\cdot) + cq(\cdot)$ (for some constant $c > 1$). The inverse f' can be chosen as a restriction of f'_1 .

Proof. Let f' be the restriction of f'_1 to the set

$$\{y \in \text{Dom}(f'_1) : |y| \leq q(|f'_1(y)|) \text{ and } |f'_1(y)| \leq q(|y|)\}.$$

Then f' obviously has balance $\leq q$. Note that since f'_1 is an inverse of f we have $\text{Im}(f) \subseteq \text{Dom}(f'_1)$. To show that f' is an inverse of f it is sufficient to check that the domain of f' contains $\text{Im}(f)$. Let $y = f(x) \in \text{Im}(f)$ for some $x \in \text{Dom}(f)$. Then $f(f'_1(y)) = y$, since f'_1 is an inverse.

Checking $|y| \leq q(|f'_1(y)|)$: $|y| = |f(f'_1(y))| \leq q(|f'_1(y)|)$; the inequality holds since q is a balance for f on input $f'_1(y)$.

Checking $|f'_1(y)| \leq q(|y|)$: $|f'_1(y)| \leq q(|f(f'_1(y))|)$ since q is a balance for f on input $f'_1(y)$; and $q(|f(f'_1(y))|) = q(|y|)$ since $f(f'_1(y)) = y$.

To find a time-complexity bound for f' , we first compute $f'_1(y)$ in time $\leq T(|y|)$; thereby we also verify that $y \in \text{Dom}(f'_1)$. To check whether y is in the domain of f' we first compare $|y|$ and $|f'_1(y)|$ in time $\leq |y| + 1$.

Checking $|y| \leq q(|f'_1(y)|)$: If $|y| \leq |f'_1(y)|$ then we automatically have $|y| \leq q(|f'_1(y)|)$. If $|y| \geq |f'_1(y)|$ we compute $q(|f'_1(y)|)$ in time $O(q(|f'_1(y)|)) (\leq O(q(|y|)))$, by writing the number $|f'_1(y)|$ in binary, and then evaluating q (see Section 4 of [2] for a similar computation). Then we check $|y| \leq q(|f'_1(y)|)$ in time $\leq |y| + 1$. Checking $|f'_1(y)| \leq q(|y|)$ is done in a similar way, in time $\leq O(q(|y|)) + |y| + 1$. \square

Theorem 2.8 *The map $\text{evR}_{q_2}^{CC}$ is complete for \mathcal{RM}_2^P with respect to inversive Turing reduction.*

Proof. Lemma 2.1 provides the following simulation of ϕ_w by $\text{evR}_{q_2}^{CC}$:

$$\phi_w = \text{decode} \circ \rho_{\text{code}(w')}_{11} \circ \text{contr} \circ \text{recontr}^{2m} \circ \text{evR}_{q_2}^{CC} \circ \text{reexpand}^m \circ \text{expand} \circ \gamma_w^o.$$

To obtain an inversive Turing simulation, let e' be any inverse of $\text{evR}_{q_2}^{CC}$. Slightly modifying the proof of Prop. 5.6 in [2], we apply e' to any string of the form

$$\text{code}(\text{ex}(w)) \ 11 \ 0^{N_{m+1}} \ 11 \ \text{code}(p) \ 11 \ z,$$

where $p \in \phi_w(\text{domC}(\phi_w))$, and $z \in A^*$; then for any $p \in \text{imC}(\phi_w) (\subseteq \phi_w(\text{domC}(\phi_w)))$, and $z \in A^*$:

$$\begin{aligned} e'(\text{code}(\text{ex}(w)) \ 11 \ 0^{N_{m+1}} \ 11 \ \text{code}(p) \ 11 \ z) \\ = \text{code}(\text{ex}(w)) \ 11 \ 0^{N_{m+1}} \ 11 \ \text{code}(t) \ 11 \ z, \end{aligned}$$

for some $t \in \phi_w^{-1}(p) \subseteq \text{Dom}(\phi_w)$. Based on e' we now construct an inverse ϕ'_w of ϕ_w such that $\phi'_w \preceq_{\text{T}} e'$; for any $y \in \text{Im}(\phi_w)$ we define

$$\phi'_w(y) = \text{decode} \circ \rho_{\text{code}(w')}_{11} \circ \text{contr} \circ \text{recontr}^{2m} \circ e' \circ \text{reexpand}^m \circ \text{expand} \circ \delta_w^o(y).$$

Here, $\delta_w^o(y)$ is defined by

$$\delta_w^o(y) = \text{code}(w) \ 11 \ \text{code}(p) \ 11 \ z,$$

when $y = pz$ with $p \in \text{imC}(\phi_w)$, $z \in A^*$. So, $\delta_w^o(\cdot)$ is similar to $\gamma_w^o(\cdot)$, except that $\delta_w^o(\cdot)$ uses $\text{imC}(\phi_w)$, whereas $\gamma_w^o(\cdot)$ uses $\text{domC}(\phi_w)$. We saw that $\gamma_w^o \in \mathcal{RM}_2^P$; but unless $P = NP$, δ_w^o will not be in \mathcal{RM}_2^P in general.

The value $\delta_w^o(y)$ can be computed by an \mathcal{RM}_2^P -machine M that makes oracle calls to $\text{Dom}(e')$ and to e' as follows. On input y , M considers all prefixes of y of increasing lengths, p_1, \dots, p_k , until $p_j \in \text{Im}(\phi_w)$ is found. Since p_j is the first prefix in $\text{Im}(\phi_w)$, we have $p_j \in \text{imC}(\phi_w)$ and $\delta_w^o(y) = \text{code}(w) \ 11 \ \text{code}(p_j) \ 11 \ z$. To test for each p_i whether $p_i \in \text{Im}(\phi_w)$, M pads p_i to produce $0^{N_{m+1}} \ 11 \ \text{code}(p_i)$; if $p_i \in \text{Im}(\phi_w)$ then $e'(\text{code}(\text{ex}(w)) \ 11 \ \bullet)$ is defined on input $0^{N_{m+1}} \ 11 \ \text{code}(p_i)$. Thus, if $\text{code}(\text{ex}(w)) \ 11 \ 0^{N_{m+1}} \ 11 \ \text{code}(p_i) \notin \text{Dom}(e')$, then $p_i \notin \text{Im}(\phi_w)$. On the other hand, if $\text{code}(\text{ex}(w)) \ 11 \ 0^{N_{m+1}} \ 11 \ \text{code}(p_i) \in \text{Dom}(e')$, then let $t_i \in \phi_w^{-1}(p_i)$ be such that

$$e'(\text{code}(\text{ex}(w)) \ 11 \ 0^{N_{m+1}} \ 11 \ \text{code}(p_i)) = \text{code}(\text{ex}(w)) \ 11 \ 0^{N_{m+1}} \ 11 \ \text{code}(t_i).$$

One oracle call to e' yields this, and hence t_i . Then we can use ϕ_w to check whether $t_i \in \text{Dom}(\phi_w)$; and this holds iff $p_i \in \text{Im}(\phi_w)$. This way, M can check whether $p_i \in \text{Im}(\phi_w)$. Thus, if $y \in \text{Im}(\phi_w)$, M will find $p_j \in \text{Im}(\phi_w)$. When $y \notin \text{Im}(\phi_w)$, M produces no output; this doesn't matter since we do not care how ϕ'_w is defined outside of $\text{Im}(\phi_w)$.

Once $\delta_w^o(y)$ is known, the remaining simulation

$$\text{decode} \circ \rho_{\text{code}(w')}_{11} \circ \text{contr} \circ \text{recontr}^{2m} \circ e' \circ \text{reexpand}^m \circ \text{expand}$$

of e' , applied to $\delta_w^o(y) = \text{code}(w) \ 11 \ \text{code}(p) \ 11 \ z$, yields $\phi'_w(y)$.

The function ϕ'_w is an inverse of ϕ_w : Indeed, for $x \in \text{Dom}(\phi_w)$, we have $\phi_w(x) = pz$ for some $p \in \text{imC}(\phi_w)$, $z \in A^*$. Then

$$\text{reexpand}^m \circ \text{expand} \circ \delta_w^o(pz) = \text{code}(\text{ex}(w)) \ 11 \ 0^{N_{m+1}} \ 11 \ \text{code}(p) \ 11 \ z;$$

and applying e' then yields

$$\text{code}(\text{ex}(w)) \ 11 \ 0^{N_{m+1}} \ 11 \ \text{code}(t) \ 11 \ z,$$

for some $t \in \phi_w^{-1}(p)$. Applying

$$\text{decode} \circ \rho_{\text{code}(w')} 11 \circ \text{contr} \circ \text{recontr}^{2m}$$

now yields tz . Finally, $\phi_w(tz) = pz$, since $t \in \phi_w^{-1}(p)$. So, $\phi_w \phi'_w \phi_w(x) = \phi_w \phi'_w(pz) = \phi_w(tz) = pz = \phi_w(x)$. \square

We show next that $\text{evR}_{q_2}^{CC}$ is not only complete for \mathcal{RM}_2^P , but for all of fP .

Proposition 2.9 *The map $\text{evR}_{q_2}^{CC} (\in \mathcal{RM}_2^P)$ is complete for fP with respect to $\leq_{\text{inv}, \top}$.*

Proof. By Prop. 5.6 in [2], $\text{ev}_{q_2}^C$ is complete in fP for inversive simulation. By Prop. 5.17 in [2], $\text{ev}_{q_2}^C \leq_{\text{inv}} (\text{ev}_{q_2}^C)^C$. Moreover, $(\text{ev}_{q_2}^C)^C \leq_{\text{inv}, \top} \text{evR}_{q_2}^{CC}$; indeed, $(\text{ev}_{q_2}^C)^C \in \mathcal{RM}_2^P$ (since $f \mapsto f^C$ maps into \mathcal{RM}_2^P), and we just saw that $\text{evR}_{q_2}^{CC}$ is complete in \mathcal{RM}_2^P . Hence $\text{ev}_{q_2}^C \leq_{\text{inv}} (\text{ev}_{q_2}^C)^C \leq_{\text{inv}, \top} \text{evR}_{q_2}^{CC}$. \square

3 Non-finite generation

In [2] we proved that fP is finitely generated, and we left open the question whether \mathcal{RM}_2^P is also finitely generated. We will now answer this question negatively. We will use the following general compactness property: If a semigroup S is finitely generated, and if Γ is any infinite generating set of S , then S is generated by some finite subset of this set Γ .

Theorem 3.1. *\mathcal{RM}_2^P is not finitely generated.*

Proof. We saw that \mathcal{RM}_2^P is generated by the infinite set

$$\begin{aligned} & \{\rho_0, \rho_1, \pi_0, \pi_1, \text{decode}_2, \text{contr}, \text{recontr}, \text{evR}_{q_2}^{CC}, \text{reexpand}, \text{expand}\} \\ & \cup \{\gamma_w : w \text{ is an } \mathcal{RM}_2^P\text{-program}\}. \end{aligned}$$

Let us assume, by contradiction, that \mathcal{RM}_2^P is finitely generated. Then a finite generating set can be extracted from this infinite generating set, so \mathcal{RM}_2^P is generated by

$$\Gamma_{\text{fin}} = \{\rho_0, \rho_1, \pi_0, \pi_1, \text{decode}_2, \text{contr}, \text{recontr}, \text{evR}_{q_2}^{CC}, \text{reexpand}, \text{expand}\} \cup \{\gamma_i : i \in F\},$$

where F is some finite set of \mathcal{RM}_2^P -programs. So for every γ_w there is a word in Γ_{fin}^* that expresses γ_w as a finite sequence of generators. Recall that $\text{Dom}(\gamma_w) = \text{code}(w) 11 \text{Dom}(\phi_w)$, and for any $x \in \text{Dom}(\phi_w)$,

$$\gamma_w(\text{code}(w) 11 x) = \text{code}(w) 11 \text{code}(u) 11 v,$$

where $x = uv$ and $u \in \text{domC}(\phi_w)$.

The proof strategy will consist in showing that there are infinitely many functions γ_w that do not have a correct representation over Γ_{fin} . More precisely, for all \mathcal{RM}_2^P -programs w and all $u \in \text{domC}(\phi_w)$, we have $\gamma_w(\text{code}(w) 11 u) = \text{code}(w) 11 \text{code}(u) 11$; so $\gamma_w(\text{code}(w) 11 u) \in \{00, 01\}^* 11 \{00, 01\}^* 11$. On the other hand, we will show that there exist (infinitely many) \mathcal{RM}_2^P -programs w such that for every $X \in (\Gamma_{\text{fin}})^*$ that represents γ_w , there exist (infinitely many) $u \in \text{domC}(\phi_w)$ such that: $X(\text{code}(w) 11 u) = \text{code}(w) 11 \text{code}(u_1) 11 u_2$, where u_2 is non-empty; so, $X(\text{code}(w) 11 u) \notin \{00, 01\}^* 11 \{00, 01\}^* 11$. Thus we obtain a contradiction.

We consider the \mathcal{RM}_2^P -programs w such that $\text{domC}(\phi_w)$ satisfies:

- (1) no word in $\text{domC}(\phi_w)$ contains 11 as a subsegment;
- (2) for all $i \in F$, $\text{domC}(\phi_i) \neq \text{domC}(\phi_w)$;
- (3) for any $u \in \text{domC}(\phi_w)$ and any integer $n > 0$, there exists $v \in \text{domC}(\phi_w)$ of length $|v| > n$ such that $u = u_0c$, u_0 is a prefix of v , and $|c| \leq 4$. Equivalently:
 $(\forall u \in \text{domC}(\phi_w)) (\forall n > 0) (\exists v \in \text{domC}(\phi_w), |v| > n) (\exists u_0, c, z \in A^*) [v = u_0z, u = u_0c, |c| \leq 4].$

We can picture this as a path in the tree of A^* , labeled by u and ending at vertex u ; at vertex u_0 along this path, at distance ≤ 4 from vertex u , a second path branches off and ends at vertex v (of length $|v| > n$).

The following family of examples shows that there exist infinitely many \mathcal{RM}_2^P -programs w that satisfy properties (1)-(3). In each of these examples (parameterized by $a \in \{0,1\}^*$) we have

$$\text{domC}(\phi_w) = \{\text{code}(a^n) 0010 : n > 0\},$$

where $a \in \{0,1\}^*$ is any fixed non-empty word (depending on w), chosen so that $\text{domC}(\phi_w) \neq \text{domC}(\phi_i)$ for all $i \in F$ (thus property (2) holds). Any word a that is long enough will work; indeed, for different words a the above prefix codes are different, whereas F is finite. Property (1) follows from the definition of code (namely, $\text{code}(0) = 00$, $\text{code}(1) = 01$). Property (3) holds because for every $u = \text{code}(a^m) 0010$ and every $n > 0$, we can take $u_0 = \text{code}(a^m)$ and $v = \text{code}(a^{n+m}) 0010$. The set $\{\text{code}(a^n) 0010 : n > 0\}$ is a regular language, with regular expression $(\text{code}(a))^+ 0010$.

Let $X \in \Gamma_{\text{fin}}^*$ be a representation of γ_w , where w is any \mathcal{RM}_2^P -program from the family of examples above with properties (1)-(3). We will consider certain suffixes S_i of X , over Γ_{fin} .

Let S_0 be the shortest suffix of X such that for all $u \in \text{domC}(\phi_w)$, $S_0(\text{code}(w) 11 u)$ is of the form $\text{code}(x) 11 \text{code}(y) 11 z \in \{00,01\}^* 11 \{00,01\}^* 11 \{0,1\}^*$. Then S_0 exists since X itself (representing γ_w) maps $\text{code}(w) 11 u$ to $\text{code}(w) 11 \text{code}(u) 11 \in \{00,01\}^* 11 \{00,01\}^* 11 \{0,1\}^*$.

Inductively we define $S_1, S_2, \dots, S_i, \dots$, where S_i is the shortest suffix of X that has S_{i-1} as a strict suffix, and such that for all $u \in \text{domC}(\phi_w)$ we have:

$$S_i(\text{code}(w) 11 u) \in \{00,01\}^* 11 \{00,01\}^* 11 \{0,1\}^*.$$

So, $S_i(\text{code}(w) 11 u)$ is of the form $\text{code}(w_1) 11 \text{code}(u_1) 11 u_2$ with $w_1, u_1, u_2 \in A^*$. Then $X = S_N$ for some $N \geq 0$ (and $|X| > N$).

Theorem 3.1 now follows from the next Lemma, according to which there are (infinitely many) $u \in \text{domC}(\phi_w)$ such that $S_N(\text{code}(w) 11 u) = \text{code}(w) 11 \text{code}(u_1) 11 u_2$, with u_2 non-empty. On the other hand, $X = S_N$, and X represents γ_w , hence by the definition of γ_w we have for every $u \in \text{domC}(\phi_w)$: $S_N(\text{code}(w) 11 u) = \text{code}(w) 11 \text{code}(u) 11$; so u_2 is empty. Thus, the assumption that X (over the finite generating set Γ_{fin}) represents γ_w , leads to a contradiction.

Lemma 3.2 *Let γ_w be such that $\text{domC}(\phi_w) = \{\text{code}(a^n) 0010 : n > 0\}$ for some word $a \in \{0,1\}^*$, chosen so that the program w satisfies properties (1)-(3). Let X be a word over Γ_{fin} that represents γ_w , and let $|X|$ be the length of X over Γ_{fin} . Let S_0, \dots, S_N be the suffixes of X defined above, with $S_N = X$. Then there exist ℓ and n with $\ell > n > 0$ such that for all $i = 0, \dots, N$ and all $u \in \text{domC}(\phi_w)$ with $|u| \geq \ell$:*

$$S_i(\text{code}(w) 11 u) = \text{code}(w_1) 11 \text{code}(u_1) 11 u_2,$$

for some $w_1, u_1, u_2 \in A^$. Moreover, u_2 has a non-empty common suffix with u , and this common suffix has length at least n .*

Proof. We have for all $u \in \text{domC}(\phi_w)$: $S_i(\text{code}(w) 11 u) = \text{code}(w_1) 11 \text{code}(u_1) 11 u_2$, for some $w_1, u_1, u_2 \in A^*$. We want to show that there is ℓ such that for all $u \in \text{domC}(\phi_w)$ with $|u| \geq \ell$: u_2 has a non-empty (sufficiently long) suffix in common with u ; the number n is an auxiliary parameter. We take u of the form $u = \text{code}(a^m) 0010$ and use induction on $i = 0, \dots, N$.

Proof for S_0 : The only generators from Γ_{fin} that can occur in S_0 are $\pi_0, \pi_1, \rho_0, \rho_1$ and γ_j (for $j \in F$). Indeed, the other generators in Γ_{fin} (namely decode_2 , contr , recontr , $\text{evR}_{q_2}^{CC}$, reexpand , expand) are only applicable to inputs of the form $\text{code}(x) 11 \text{code}(y) 11 z$; so, S_0 would end before a generator in $\{\text{decode}_2, \text{contr}, \text{recontr}, \text{evR}_{q_2}^{CC}, \text{reexpand}, \text{expand}\}$ can be applied. Moreover, S_0 cannot start with a generator in $\{\text{decode}_2, \text{contr}, \text{recontr}, \text{evR}_{q_2}^{CC}, \text{reexpand}, \text{expand}\}$; indeed, for all inputs $\text{code}(w) 11 u \in \text{domC}(X)$, $u = \text{code}(a^m) 0010$ contains no 11, so these generators are not defined on any element of $\text{domC}(X)$. So, S_0 is over $\{\pi_0, \pi_1, \rho_0, \rho_1\} \cup \{\gamma_j : j \in F\}$.

The actions of $\pi_0, \pi_1, \rho_0, \rho_1$ can change an input in at most $|S_0|$ positions at the left end of the input, so these actions preserve a common suffix u_2 and u of length $\geq |u| - |S_0|$. Thus, if S_0 consists only of instances of $\pi_0, \pi_1, \rho_0, \rho_1$, the Lemma holds for S_0 if $|u| \geq \ell \geq n + |S_0|$ and $n > 0$.

Suppose now that S_0 contains γ_j for some $j \in F$. Then (if $m > |S_0|/|a|$), instances of $\pi_0, \pi_1, \rho_0, \rho_1$ will transform the input $u = \text{code}(w) 11 \text{code}(a^m) 0010$ into a word $\text{code}(x) 11 s \text{code}(a^k) 0010$ (for some $x, s \in A^*$, $k > 0$), such that γ_j can be applied. This action changes an input in $< |S_0|$ positions at the left end of the input. Since γ_j is assumed to be applicable now, we must also have $x = j$ and $s = y_0 z$ for some $y_0 \in \text{domC}(\phi_j)$, $z \in A^*$. Then the output of γ_j is $\gamma_j(\text{code}(j) 11 s \text{code}(a^k) 0010) = \text{code}(j) 11 \text{code}(y_0) 11 z \text{code}(a^k) 0010$, thus the common suffix of u_2 and u could decrease by length $\leq |y_0|$ under the action of γ_j . So we let $\ell \geq n + |S_0| + |y_0|$ and $n > 0$. Also, at most one γ_j (with $j \in F$) occurs in S_0 , since after γ_j the output is of the form $\text{code}(w_1) 11 \text{code}(u_1) 11 u_2$, which marks the end of the action of S_0 . This proves the Lemma for S_0 .

Inductive step “ $S_i \rightarrow S_{i+1}$ ”, for $0 \leq i < N$: By induction we assume that for all $u \in \text{domC}(\phi_w)$ with $|u| \geq \ell$, we have $S_i(\text{code}(w) 11 u) = \text{code}(w_1) 11 \text{code}(u_1) 11 u_2$ for some $w_1, u_1, u_2 \in A^*$, where u_2 and u have a common suffix of length $\geq n$ (> 0). Let us write $S_{i+1} = T_{i+1} S_i$; then T_{i+1} is non-empty (by the definition of S_{i+1}). We also let $T_0 = S_0$.

Claim 1: If T_{i+1} contains a generator $g \in \{\text{contr}, \text{recontr}, \text{evR}_{q_2}^{CC}, \text{reexpand}, \text{expand}, \text{decode}_2\}$, then g is the first (i.e., rightmost) letter of T_{i+1} , and g occurs only once.

Indeed, if g were applicable later in T_{i+1} , the output of the generator preceding g would be of the form $\text{code}(w_1) 11 \text{code}(u_1) 11 u_2$, so S_{i+1} would have ended before g was applied.

Claim 2: If T_{i+1} contains a generator $g \in \{\text{contr}, \text{recontr}, \text{evR}_{q_2}^{CC}, \text{reexpand}, \text{expand}\} \cup \{\gamma_j : j \in F\}$, then g is the last (i.e., leftmost) letter of T_{i+1} , and g occurs only once.

Indeed, such a generator outputs a word of the form $\text{code}(w_1) 11 \text{code}(u_1) 11 u_2$. So, S_{i+1} ends after such a generator.

As a consequence of Claims 1 and 2, if T_{i+1} contains a generator $g \in \{\text{contr}, \text{recontr}, \text{evR}_{q_2}^{CC}, \text{reexpand}, \text{expand}\}$, then T_{i+1} consists of just g . A generator of this form does not change u_2 .

So we can assume for the remaining cases that T_{i+1} is of the form t_{i+1} , or $t_{i+1} \cdot \text{decode}_2$, or $\gamma_j \cdot t_{i+1}$, or $\gamma_j \cdot t_{i+1} \cdot \text{decode}_2$, where $j \in F$ and t_{i+1} is over the generators $\pi_0, \pi_1, \rho_0, \rho_1$.

Let $\text{code}(w_1) 11 \text{code}(u_1) 11 u_2$ be the input of T_{i+1} (and this is also the output of S_i), where u_2 and u have a common suffix of length $\geq n$.

- Case where T_{i+1} is over the generators $\pi_0, \pi_1, \rho_0, \rho_1$: Then T_{i+1} changes the input in at most $|T_{i+1}|$ positions at the left end of the input, so u_2 will not be affected if $\ell - n \geq |T_{i+1}|$ (and $n > 0$).
- Case where $T_{i+1} = t_{i+1} \cdot \text{decode}_2$, with t_{i+1} over $\pi_0, \pi_1, \rho_0, \rho_1$: The output of decode_2 is of the form $\text{code}(w_1) 11 u_1 u_2$, so the common suffix of u_2 and u is preserved by decode_2 . The action of t_{i+1} , containing only generators from $\{\pi_0, \pi_1, \rho_0, \rho_1\}$, affects at most $|t_{i+1}|$ positions near the left side of the input, so u_2 is not changed if $\ell - n \geq |T_{i+1}|$ (and $n > 0$).
- Case where $T_{i+1} = \gamma_j \cdot t_{i+1}$, with t_{i+1} over $\pi_0, \pi_1, \rho_0, \rho_1$: Applications of $\pi_0, \pi_1, \rho_0, \rho_1$ change fewer than $|t_{i+1}|$ letters of the input near the left end, so the common suffix is not affected if $\ell - n \geq |T_{i+1}|$. When γ_j is applied, the output produced will be of the form $\text{code}(j) 11 \text{code}(y_{i+1}) 11 z \text{code}(a^n) 0010$, where $y_{i+1} \in \text{domC}(\phi_j)$. Then u_2 will not be affected if we pick $\ell \geq n + |T_{i+1}| + |y_{i+1}|$ and $n > 0$.
- Case where $T_{i+1} = \gamma_j \cdot t_{i+1} \cdot \text{decode}_2$, with t_{i+1} over $\pi_0, \pi_1, \rho_0, \rho_1$: This case can be handled as a combination of the previous two cases.

In all the above cases the constraints are fulfilled for all $i = 0, \dots, N$, and for all $u = \text{code}(a^m) 0010$, if $m \geq N + |X| + \sum_{i=0}^N |y_i|$ (using the fact that $\sum_{i=0}^N |T_i| = |X|$). Note that the words y_i do not depend on the choice of the input $u = \text{code}(a^m) 0010$, whenever m is long enough; indeed, to determine all y_i we can apply each S_i to the infinite word $\text{code}(a)^\omega \in \{0, 1\}^\omega$. \square

Notation. For a given polynomial q (of the form $q(n) = a n^k + a$ with integers $a, k \geq 1$), let

$$\mathcal{S}_2^{(q)} = \{f \in \mathcal{RM}_2^P : f \text{ is computed by an } \mathcal{RM}_2^P\text{-program with built-in polynomial } \leq q\}.$$

We call w an $\mathcal{S}_2^{(q)}$ -program iff w is an \mathcal{RM}_2^P -program with built-in balance and time-complexity polynomial $\leq q$.

Let $\mathcal{RM}_2^{(q)} = \langle \mathcal{S}_2^{(q)} \rangle$, i.e., the submonoid of \mathcal{RM}_2^P generated by the set $\mathcal{S}_2^{(q)}$. Obviously, we have:

Proposition 3.3 *For any set of polynomials $\{q_i : i \in \mathbb{N}\}$ of the form $q_i(n) = a_i n^{k_i} + a_i$, such that $\sup\{a_i : i \in \mathbb{N}\} = +\infty = \sup\{k_i : i \in \mathbb{N}\}$, we have: $\bigcup_{i \in \mathbb{N}} \mathcal{RM}_2^{(q_i)} = \mathcal{RM}_2^P$. \square*

The non-finite generation result for \mathcal{RM}_2^P also holds for $\mathcal{RM}_2^{(q)}$, and the proof is similar. We need a few preliminary facts.

Lemma 3.4 *For every polynomial q of the form $q(n) = a n^k + a$ with $a, k \geq 2$, and every $\mathcal{S}_2^{(q)}$ -program w we have: $\gamma_w \in \mathcal{S}_2^{(q)}$.*

Proof. Recall that $\gamma_w(\text{code}(w) 11 uv) = \text{code}(w) 11 \text{code}(u) 11 v$, where $u \in \text{domC}(\phi_w)$. The input balance of γ_w is $\leq q$. Indeed, the input is shorter than the output; and the output length is $2|w| + 2 + 2|u| + 2 + |v|$, which is less than $q(|\text{code}(w) 11 uv|) = q(2|w| + 2 + |u| + 2 + |v|)$ when $q(n) \geq 2n^2 + 2$.

To compute $\text{code}(w) 11 \text{code}(u) 11 v$ from input $\text{code}(w) 11 uv$, an \mathcal{RM}_2^P -machine can proceed as follows: First, the machine reads and outputs $\text{code}(w) 11$. Then it runs the program w on input uv , i.e., it simulates the corresponding \mathcal{RM}_2^P -machine M_w (which has built-in polynomial q), with an extra tape and a few modifications. While searching for a prefix of uv in $\text{domC}(\phi_w)$, the longest prefix examined so far is kept on the extra tape; the output $\phi_w(u)$ of M_w will not be written on the output tape. Once u (the prefix of uv in $\text{domC}(\phi_w)$) has been found (and written on the extra tape), $\text{code}(u) 11 v$ is appended on the output tape.

All this takes time $\leq |\text{code}(w) 11| + q(|u|) + |\text{code}(u) 11 v| = 2|w| + 2 + q(|u|) + 2|u| + 2 + |v|$; this is $< q(2|w| + 2 + |u| + 2 + |v|) = q(|\text{code}(w) 11 uv|)$ when $q(n) \geq 2n^2 + 2$. \square

Lemma 3.5 *Let q be a polynomial that is larger than a certain polynomial of degree 5. Then $\mathcal{RM}_2^{(q)}$ is generated by*

$$\{\rho_0, \rho_1, \pi_0, \pi_1, \text{decode}_2, \text{contr}, \text{recontr}, \text{evR}_{q_2}^{CC}, \text{reexpand}, \text{expand}\} \cup \{\gamma_z : z \text{ is an } \mathcal{S}_2^{(q)}\text{-program}\}.$$

Proof. When w is an $\mathcal{S}_2^{(q)}$ -program then as a consequence of Lemma 2.1,

$$\begin{aligned} \phi_w &= \rho_{\text{code}(w')} 11 \circ \text{decode}_2 \circ \text{contr} \circ \text{recontr}^{2m} \circ \text{evR}_{q_2}^{CC} \circ \text{reexpand}^m \circ \text{expand} \circ \gamma_w^o, \\ &= \rho_{\text{code}(w')} 11 \circ \text{decode}_2 \circ \text{contr} \circ \text{recontr}^{2m} \circ \text{evR}_{q_2}^{CC} \circ \text{reexpand}^m \circ \text{expand} \circ \gamma_w \circ \pi_{\text{code}(w) 11}, \end{aligned}$$

where q_2 is a certain polynomial of degree 2. So the above generating set does indeed generate $\mathcal{RM}_2^{(q)}$. We still need to show that these generators belong to $\mathcal{RM}_2^{(q)}$.

The functions $\rho_0, \rho_1, \text{decode}_2, \text{contr}, \text{recontr}, \text{reexpand}, \text{expand}, \pi_0, \pi_1$ have balance and complexity $\leq 4(n+1)^2$. And $\gamma_w \in \mathcal{RM}_2^{(q)}$ if w is an $\mathcal{S}_2^{(q)}$ -program (by Lemma 3.4). Let us verify that $\text{evR}_{q_2}^{CC}$ has balance $\leq q_2$ and complexity $O(n^5)$. By definition,

$$\text{evR}_{q_2}^{CC}(\text{code}(w) 11 \text{code}(u) 11 v) = \text{code}(w) 11 \text{code}(\phi_w(u)) 11 v.$$

Then $\text{evR}_{q_2}^{CC}$ has balance $\leq q_2$, since on an output of length $n = 2|w| + 2 + 2|\phi_w(u)| + 2$, the input length is $\leq 2|w| + 2 + 2q_2(|\phi_w(u)|) + 2 \leq q_2(2|w| + 2 + 2|\phi_w(u)| + 2) = q_2(n)$.

When ϕ_w can be computed by an $\mathcal{RM}_2^{(q)}$ -machine with built-in polynomial p_w ($\leq q_2$), then $\text{evR}_{q_2}^{CC}(\text{code}(w) 11 \text{code}(u) 11 v)$ can be computed in time $\leq c|w| p_w(|u|)^2 \leq c|w| q_2(|u|)^2$, for some constant $c > 0$ (see the proof of Prop. 4.4 in [2]). Since q_2 has degree 2, $\text{evR}_{q_2}^{CC}$ has complexity $O(n^5)$. Thus, there exists q of degree 5 such that the above generators belong to $\mathcal{RM}_2^{(q)}$. \square

Theorem 3.6 *For any polynomial q such that $q(n) = an^k + a$, with $k \geq 5$ and $a > a_0$ (for some constant $a_0 > 1$), we have: $\mathcal{RM}_2^{(q)}$ is not finitely generated.*

Proof. The proof is very similar to the proof of Theorem 3.1. We saw in Lemma 3.5 that $\mathcal{RM}_2^{(q)}$ is generated by the infinite set

$$\{\rho_0, \rho_1, \pi_0, \pi_1, \text{decode}_2, \text{contr}, \text{recontr}, \text{evR}_{q_2}^{CC}, \text{reexpand}, \text{expand}\} \cup \{\gamma_z : z \text{ is an } \mathcal{S}_2^{(q)}\text{-program}\}.$$

Let us assume, by contradiction, that $\mathcal{RM}_2^{(q)}$ is finitely generated. Then a finite generating set can be extracted from this infinite generating set; so $\mathcal{RM}_2^{(q)}$ is generated by

$$\Gamma_{\text{fin}} = \{\rho_0, \rho_1, \pi_0, \pi_1, \text{decode}_2, \text{contr}, \text{recontr}, \text{evR}_{q_2}^{CC}, \text{reexpand}, \text{expand}\} \cup \{\gamma_i : i \in F\},$$

where F is some finite set of $\mathcal{S}_2^{(q)}$ -programs. For every $\mathcal{S}_2^{(q)}$ -program w let X be a word in Γ_{fin}^* that expresses γ_w as a finite sequence of generators.

From here on, the proof is identical to the proof of Theorem 3.1. We use the fact that $\text{domC}(\phi_w) = \{\text{code}(a^n) 0010 : n > 0\}$ is a finite-state language, so for such a program w , γ_w has linear complexity (being computable by a Mealy machine) and belongs to $\mathcal{S}_2^{(q)}$. \square

4 Some complexity consequences of non-finite generation

4.1 Hierarchy and separation

Proposition 4.1 *Let q be a polynomial of the form $q(n) = an^k + a$ such that $a, k \geq 1$. The set $\mathcal{S}_2^{(q)}$, and hence the monoid $\mathcal{RM}_2^{(q)}$, are contained in a finitely generated submonoid of \mathcal{RM}_2^P .*

Proof. Let w be a \mathcal{RM}_2^P -program such that ϕ_w has I/O-balance and time-complexity $\leq q$. Then evR_q^C can simulate ϕ_w directly, without any need of padding and unpadding. So we have for all $u \in \text{domC}(\phi_w)$, $v \in A^*$:

$$\phi_w(uv) = \rho_{\text{code}(w)11} \circ \text{evR}_q^C \circ \pi_{\text{code}(w)11}(uv).$$

So $\mathcal{S}_2^{(q)}$ is contained in the submonoid generated by $\{\pi_0, \pi_1, \rho_0, \rho_1, \text{evR}_q^C\}$. (Compare with Lemma 2.2 and the proof of Prop. 4.5 in [2].) \square

The proof of Prop. 4.1 yields the following chain of submonoids in which non-finitely generated and finitely generated submonoids alternate.

Corollary 4.2 *Let $\dots < q_i < q_{i+1} < \dots$ be any sequence of polynomials such that for all $i \geq 0$, q_{i+1} is large enough so that $\text{evR}_{q_i}^C$ has an \mathcal{RM}_2^P -program with built-in polynomial q_{i+1} . Then \mathcal{RM}_2^P contains a strict inclusion chain, which is infinite in the upward direction,*

$$\dots \subsetneq \mathcal{RM}_2^{(q_i)} \subsetneq \langle \pi_0, \pi_1, \rho_0, \rho_1, \text{evR}_{q_i}^C \rangle_{\mathcal{RM}_2^P} \subsetneq \mathcal{RM}_2^{(q_{i+1})} \subsetneq \dots \dots$$

Proof. The strictness of the inclusions in the chain follows from the fact that non-finite generation and finite generation alternate. \square

Theorem 4.3 *Let q be a polynomial of the form $q(n) = an^k + a$ such that $a > 1, k \geq 1$. The submonoid $\mathcal{RM}_2^{(q)} \subseteq \mathcal{RM}_2^P$ has the following properties:*

- (1) $\mathcal{RM}_2^{(q)} \neq \mathcal{RM}_2^P$.
- (2) If $q(n) \geq 2(n+1)^2$ (for all $n \in \mathbb{N}$), then $\mathcal{RM}_2^{(q)}$ contains elements of arbitrarily high polynomial balance and time-complexity.
- (3) $\mathcal{S}_2^{(q)} \neq \mathcal{RM}_2^{(q)}$, if $q(n) \geq 2(n+1)^2$.

(4) $\text{evR}_q^C \notin \mathcal{RM}_2^{(q)}$, if $k \geq 5$ and $a \geq a_0$ (where a_0 is as in Theorem 3.6).

Moreover, evR_q^C has balance $\leq q$, but its time-complexity is not $\leq q$.

(5) Let q_1, q_2 be polynomials of the above form, such that $q_1(n) < q_2(n)$ for all $n \in \mathbb{N}$. Suppose also that $q_1(n) = a n^k + a$ with $k \geq 5$ and $a \geq a_0$ (as in (4)), and that q_2 is large enough so that $\text{evR}_{q_1}^C \in \mathcal{RM}_2^{(q_2)}$. Then $\mathcal{RM}_2^{(q_1)} \subsetneq \mathcal{RM}_2^{(q_2)}$.

Proof. (1) Since $\mathcal{RM}_2^{(q)}$ is contained in a finitely generated submonoid of \mathcal{RM}_2^P (Prop. 4.1), and \mathcal{RM}_2^P is not contained (by \subseteq) in a finitely generated submonoid of \mathcal{RM}_2^P (itself), inequality follows.

(2) Consider the function $s : 0^n 1 x \mapsto 0^{2n^2} 1 x$, for all $n \geq 0, x \in \{0, 1\}^*$. Then s has time-complexity $\leq 2(n+1)^2$. Indeed, a Turing machine on input $0^n 1$ can read this word n times, each time turning an input 0 into some new letter a , and each time writing 0^n on the output tape; this produces 0^{n^2} in the output; then one more copy of 0^{n^2} is made, followed by 1. This takes time $\leq 2(n+1)^2$.

Then, s^m (i.e., the composition of m instances of s) has complexity $\geq 2^m n^{2^m}$ (since the output length is that high, the time must be at least that much too). Thus the functions $s^m \in \mathcal{RM}_2^{(q)}$ (as m grows) have unbounded complexity, both in their degree and in their coefficient.

(3) By (2), $\mathcal{RM}_2^{(q)}$ contains functions with arbitrarily high polynomial balance and time-complexity, whereas $\mathcal{S}_2^{(q)}$ only contains functions with balance and complexity $\leq q$.

(4) By Prop. 4.1, $\mathcal{RM}_2^{(q)}$ is contained in the submonoid generated by $\{\pi_0, \pi_1, \rho_0, \rho_1, \text{evR}_q^C\}$; and we easily see that $\pi_0, \pi_1, \rho_0, \rho_1 \in \mathcal{RM}_2^{(q)}$. Hence, if evR_q^C belonged to $\mathcal{RM}_2^{(q)}$, the monoid $\mathcal{RM}_2^{(q)}$ would be finitely generated, contradicting Theorem 3.6.

The input balance of evR_q^C is $\leq q$ (see Lemma 3.5, where this is proved for q_2). It follows that the time-complexity of evR_q^C is not $\leq q$, otherwise we would have $\text{evR}_q^C \in \mathcal{RM}_2^{(q)}$.

(5) This follows from (4) since $\text{evR}_q^C \notin \mathcal{RM}_2^{(q_1)}$, but $\text{evR}_q^C \in \mathcal{RM}_2^{(q_2)}$. \square

Corollary 4.4 (Strict complexity hierarchy of submonoids in \mathcal{RM}_2^P). *There exists an infinite sequence of polynomials $(q_i : i \in \mathbb{N})$, each of the form $q_i(n) = a_i n^{k_i} + a_i$ with $k_i, a_i > 1$, and with $q_i(n) < q_{i+1}(n)$ for all $i, n \in \mathbb{N}$, such that the following holds:*

$$\mathcal{RM}_2^{(q_i)} \subsetneq \mathcal{RM}_2^{(q_{i+1})} \text{ for all } i, \text{ and } \bigcup_{i \in \mathbb{N}} \mathcal{RM}_2^{(q_i)} = \mathcal{RM}_2^P.$$

Moreover, \mathcal{RM}_2^P (which is not finitely generated) is the union of a \subset -chain of 4-generated submonoids.

Proof. The first statements follow from Theorem 4.3 (1) and Prop. 3.3. The last statement follows from Cor. 4.2. \square

Since each $\mathcal{RM}_2^{(q_i)}$ contains functions of arbitrarily high polynomial complexity (by Theorem 4.3 (2)), the monoids $\mathcal{RM}_2^{(q_i)}$ form a strict complexity hierarchy of a new sort, different from the usual complexity hierarchies. The fact that $\mathcal{S}_2^{(q)} \neq \mathcal{RM}_2^P$ could have been shown by a diagonal argument. It is not clear whether classical separation techniques from complexity theory would show the results (1), (3), (4), (5) of Theorem 4.3.

Remark: The monoid fP , being finitely generated, does not contain an infinite strict complexity hierarchy of monoids (but it can contain hierarchies of sets). Indeed, we have in general:

Fact. *A finitely generated monoid M does not contain any infinite strict ω -chain of submonoids whose union is M .*

Indeed, if we had a chain $(M_i : i \in \omega)$ with $M_0 \subsetneq \dots \subsetneq M_i \subsetneq M_{i+1} \subsetneq \dots \subsetneq \bigcup_{i \in \omega} M_i = M$, then there would exist j such that M_j contains a finite set of generators of M (since $\bigcup_{i \in \omega} M_i = M$). Then $M_j = M$, contradicting the strict hierarchy.

This Fact does not hold for chains over arbitrary order types; it holds for limit ordinals. The non-finitely generated monoid \mathcal{RM}_2^P contains the encoding \mathbf{fP}^C as a submonoid (see Section 3 in [2]). And \mathbf{fP}^C is finitely generated (being an isomorphic copy of \mathbf{fP}), and \mathbf{fP}^C contains an isomorphic copy of \mathcal{RM}_2^P . This leads to non- ω strict chains of submonoids of \mathbf{fP} and of \mathcal{RM}_2^P .

4.2 Irreducible functions

Another consequence of non-finite generation is that \mathcal{RM}_2^P and $\mathcal{RM}_2^{(q)}$ have “irreducible” elements, i.e., elements that cannot be expressed by composition of lower-complexity elements. We make this precise in the next definitions.

In this subsection we do not use evaluation maps, so we can use “polynomials” $q(n) = an^k + a$ where we drop the requirement that a, k are integers, i.e., we now allow real numbers ≥ 1 .

Definition 4.5 *The inf complexity degree of $f \in \mathcal{RM}_2^P$ is*

$$d_f = \inf\{k \in \mathbb{R}_{\geq 1} : f \in \mathcal{S}_2^{(q)} \text{ for some polynomial } q \text{ of the form } q(n) = bn^k + b, \text{ for some } b > 1\}.$$

We also define the inf complexity coefficient c_f of f by

$$c_f = \inf\{C_f(\varepsilon) : \varepsilon \in \mathbb{R}_{>0}\}, \text{ where}$$

$$C_f(\varepsilon) = \inf\{a \in \mathbb{R}_{\geq 1} : f \in \mathcal{S}_2^{(q)} \text{ for some polynomial } q \text{ of the form } q(n) = an^{d_f+\varepsilon} + a\}.$$

The inf complexity polynomial of f is the polynomial q_f given by $q_f(n) = c_f \cdot (n^{d_f} + 1)$ (for all $n \in \mathbb{N}$).

Since d_f and c_f are defined by infimum, f might not be in $\mathcal{S}_2^{(q_f)}$. By the definition of inf we have the following.

Proposition 4.6 *For any polynomial $q(n) = an^k + a$ with $k > d_f$ and $a > c_f$: $f \in \mathcal{RM}_2^{(q)}$. \square*

On the other hand, for every $\varepsilon_1 > 0$, $\varepsilon_2 > 0$:

$$f \notin \mathcal{S}_2^{(p_1)} \text{ for any polynomial } p_1(n) = bn^{d_f-\varepsilon_1} + b \text{ with any } b > 1;$$

$$f \notin \mathcal{S}_2^{(p_2)} \text{ where } p_2(n) = (c_f - \varepsilon_2) \cdot (n^{d_f} + 1).$$

Definition 4.7 *Let us choose $\delta_1, \delta_2 \in \mathbb{R}_{>0}$. A function $f \in \mathcal{RM}_2^P$ is called (δ_1, δ_2) -reducible iff $f \in \mathcal{RM}_2^{(q)}$ for some polynomial $q(n) = (c_f - \delta_2) \cdot (n^{d_f-\delta_1} + 1)$. And f is called (δ_1, δ_2) -irreducible iff f is not (δ_1, δ_2) -reducible.*

In other words, f is (δ_1, δ_2) -reducible iff f is a *composite* of elements of $\mathcal{S}_2^{(q)}$ i.e., $f \in \mathcal{RM}_2^{(q)}$, where $q(n) = (c_f - \delta_2) \cdot (n^{d_f-\delta_1} + 1)$. So, f can be factored into functions that “have strictly lower complexity than f ” (regarding both the degree and the coefficient). Note that in the definition of d_f and c_f we used $\mathcal{S}_2^{(q)}$, not $\mathcal{RM}_2^{(q)}$ (Def. 4.5).

Proposition 4.8 *For all $\delta_1, \delta_2 \in \mathbb{R}_{>0}$ and all polynomials q_1, q_2 such that $\mathcal{RM}_2^{(q_1)} \subsetneq \mathcal{RM}_2^{(q_2)}$, there exist (δ_1, δ_2) -irreducible functions in $\mathcal{RM}_2^{(q_2)} - \mathcal{RM}_2^{(q_1)}$.*

Proof. By contradiction, assume that there exist δ_1, δ_2 such that every $f \in \mathcal{RM}_2^{(q_2)} - \mathcal{RM}_2^{(q_1)}$ is (δ_1, δ_2) -reducible, i.e., f can be factored as $f = f_m \circ \dots \circ f_1$, where $f_i \in \mathcal{RM}_2^{(q_2)}$ ($i = 1, \dots, m$) with $\inf \text{degree } d_{f_i} < d_f - \delta_1$ and $\inf \text{coefficient } c_{f_i} < c_f - \delta_2$. By the contradiction assumption, among these factors, those that are in $\mathcal{RM}_2^{(q_2)} - \mathcal{RM}_2^{(q_1)}$ can themselves be factored into elements of degree and coefficient lower by amount δ_1 , respectively δ_2 . I.e., a factor $f_i \in \mathcal{RM}_2^{(q_2)} - \mathcal{RM}_2^{(q_1)}$ can be factored as

$f_i = f_{i,m_i} \circ \dots \circ f_{i,1}$ with $d_{f_{i,j}} < d_{f_i} - \delta_1$ and $c_{f_{i,j}} < c_{f_i} - \delta_2$; hence, $d_{f_{i,j}} < d_f - 2\delta_1$ and $c_{f_{i,j}} < c_f - 2\delta_2$, for $j = 1, \dots, m_i$. By repeating this process we keep reducing the degree and the coefficient by at least δ_1 , respectively δ_2 , in each step. After a finite number of steps we obtain a factorization of f into functions in $\mathcal{RM}_2^{(q_1)}$, contradicting the assumption that $\mathcal{RM}_2^{(q_1)} \subsetneq \mathcal{RM}_2^{(q_2)}$. \square

Remark: A finitely generated monoid, like \mathbf{fP} , does not contain irreducible functions of arbitrarily large complexity. Indeed, all elements are expressible as a composite of elements of bounded complexity (namely the maximum complexity of the finitely many generators).

Acknowledgement: The paper benefitted from the referee's thoughtful reading and advice.

References

- [1] J. Berstel, D. Perrin, *Theory of Codes*, Academic Press (1985).
- [2] J.C. Birget, "Semigroups and one-way functions", *International J. of Algebra and Computation* 25.1-2 (2015) 3-36.
- [3] J.C. Birget, "Polynomial-time right-ideal morphisms and congruences", <http://arXiv.org/abs/1511.02056> (Nov. 2015).
- [4] J.C. Birget, "Monoid generalizations of the Richard Thompson groups", *J. of Pure and Applied Algebra* 213(2) (2009) 264-278.
- [5] J.C. Birget, "Circuits, coNP-completeness, and the groups of Richard Thompson", *International J. of Algebra and Computation* 16(1) (Feb. 2006) 35-90.
- [6] J.C. Birget, "The groups of Richard Thompson and complexity", *International J. of Algebra and Computation* 14(5,6) (Dec. 2004) 569-626.
- [7] J.W. Cannon, W.J. Floyd, W.R. Parry, "Introductory notes on Richard Thompson's groups", *L'Enseignement Mathématique* 42 (1996) 215-256.
- [8] W. Diffie, M. Hellman, "New directions in cryptography", *IEEE Trans. Information Theory* 22.6 (1976) 644-654.
- [9] D.Z. Du, K.I. Ko, *Theory of Computational Complexity*, Wiley (2000).
- [10] O. Goldreich, *Foundations of Cryptography: Basic Tools*, Cambridge U.P. (2001).
- [11] L.H. Hemaspaandra, M. Ogihara, *The Complexity Theory Companion*, Springer (2002).
- [12] G. Higman, "Finitely presented infinite simple groups", Notes on Pure Mathematics 8, The Australian National University, Canberra (1974).
- [13] L. Levin, "The tale of one-way functions", *Problemy Peredatshi Informatsii* 39.1 (2003) 92-103.
- [14] R. McKenzie, R.J. Thompson, "An elementary construction of unsolvable word problems in group theory", in *Word Problems*, (W. Boone, F. Cannonito, R. Lyndon, editors), North-Holland (1973) pp. 457-478.
- [15] Ch. Papadimitriou, *Computational Complexity*, Addison-Wesley (1994).
- [16] Richard J. Thompson, "Embeddings into finitely generated simple groups which preserve the word problem", in *Word Problems II*, (S. Adian, W. Boone, G. Higman, editors), North-Holland (1980) pp. 401-441.