

INPUT- OR OUTPUT-UNARY SWEEPING TRANSDUCERS ARE WEAKER THAN THEIR 2-WAY COUNTERPARTS

BRUNO GUILLON¹

Abstract. In a previous paper we showed that two-way (nondeterministic) transducers with unary input and output alphabets have the same recognition power as the sweeping ones. We show that this no longer holds when one of the alphabets has cardinality at least 2.

Mathematics Subject Classification. 68Q70.

1. INTRODUCTION

Finite automata exist in different variants: 1-way deterministic (1DFA), 1-way nondeterministic (1NFA), 2-way deterministic (2DFA) and 2-way nondeterministic (2NFA) but all are equivalent as far as recognition power is concerned. Providing an additional tape to the device leads to a new computational model. In automata theory, there are mainly two such models: two-tape finite automata and transducers. Both recognize the same object: binary word relations, *i.e.*, a subset of a direct product of two free monoids, say $\Sigma^* \times \Delta^*$. However, in the former model, the two tapes play a symmetric role, while in the latter, an input (read-only) tape and an output (write-only) tape are distinguished. In the case of 1-way devices, it is well known that both models recognize the same class of relations, namely the class of rational relations on $\Sigma^* \times \Delta^*$. But the different versions, deterministic or nondeterministic, lead to differences: 1-way deterministic transducers are less powerful than 1-way two-tape DFAs, which are themselves less powerful than 1-way transducer (or 1-way two-tape NFA), see for instance [17].

Here we work with 2-way transducers which are 2-way finite automata where every transition is associated with a regular language. Though the more classical model assumes that a transition is associated with a singleton, it is routine to prove that this more general form does not alter the computational power (indeed, regular output producing can be simulated by stationary steps simulating an automaton on the output). The machine must be thought of as consisting of a read-only input tape which can be scanned back and forth and a write-only output tape which is visited in one direction only. Every time a symbol of the input tape is read, a word from the language associated with the transition is printed on the output tape and the input head is moved in one of the two directions or is kept where it is. The accepted relation is then the set of pairs of words (u, v) such that v is produced on the output tape during an accepting computation of the automaton on u . It is a subset of $\Sigma^* \times \Delta^*$, where Σ is the input alphabet and Δ the output alphabet. A big difference from the finite automata

Keywords and phrases. Unary alphabets, Hadamard relations, two-way, sweeping transducers.

¹ LIAFA, Université Paris-Diderot, Paris 7, Paris. guillon.bruno+cs@gmail.com

TABLE 1. The most head-move restricted variant (among 1-way, sweeping and 2-way) equivalent to a 2-way transducer by cases.

Restrictions	Deterministic = [11] Functional	Nondeterministic
Input-unary	Sweeping[6]	2-way Theorem 5.3
Output-unary	1-way[2,7]	2-way Theorem 4.7
Input- and output-unary		Sweeping [7]

is that 2-way transducers, even the deterministic version, have different recognition power than their 1-way counterpart. Consequently, the issues on these objects are of a different nature and in particular it makes sense to investigate subclasses of such transducers. Without being exhaustive we recall a couple of major results in the literature. In [11] structural properties are studied: the 2-way deterministic transducers were shown to be those realizing a binary function definable in MSO. In [8] the relations realized by 2-way transducers are proved to be uniformizable. We recall that this means that for each input word with a non-empty image in the relation, it is possible to choose one and only one word in the image in such a way that the resulting function is also realizable by a 2-way transducer. Decision issues were also tackled: it is decidable whether or not given a 2-way or sweeping functional transducer is equivalent to a 1-way transducer, [3, 12].

In the majority of the results cited above, the authors consider some subfamilies of transducers where, in some sense, the nondeterminism is restricted. In [7], we studied a different kind of restriction, considering the case where both input and output alphabets are unary, *i.e.*, Σ and Δ are both singletons. Restricting the alphabets to a single letter is usual in Automata Theory. This particular case shows important differences with the general case. Probably, the main result of this kind is the collapse of the unary context-free and regular languages, provided by Ginsburg and Rice [13]. Speaking of transducers, we can deduce from [2], that whenever an output-unary 2-way transducer is functional or when its underlying automaton is unambiguous, it can be simulated by a 1-way transducer.

In [7], we introduced the Hadamard like operations on relations. As a byproduct it was shown that all unary 2-way transducers are equivalent to sweeping transducers: in other words it is no loss of generality to impose that the transducer changes direction on the endmarkers only. In this paper we investigate the cases where one only of the two alphabets is unary. The results are summarized in Table 1, in which we distinguish six cases: on one hand, a 2-way transducer may be deterministic or nondeterministic (columns); on the other hand, one or both alphabets may be unary or not (lines). In each cases, we determine the most restricted variant among 1-way, sweeping and 2-way (*i.e.*, unrestricted), which is equivalent to a general 2-way transducer. The particular and simple case of input-unary deterministic 2-way transducers has been discussed in ([6], p. 4). Our results solve the case of input- or output-unary 2-way nondeterministic transducers.

First, when the output alphabet is unary, since the image of a fixed word by a 2-way transducer is always a rational language and because unary rational languages are semi-linear sets, we study the period of such image languages. We prove that given a 1-way transducer (*resp.* a sweeping transducer), there exists a constant k such that, for each input word u , the image of u is a semi-linear language of period bounded by k (*resp.* in $\mathcal{O}(|u|^k)$, where $|u|$ denotes the length of u). Then we exhibit a relation accepted by an output-unary 2-way transducer that admits images whose periods are not polynomially bounded in the length of the input. This proves that sweeping transducers are weaker than general 2-way transducers. A recent paper shows that two-way \mathcal{N} -automata on the tropical semiring $\mathcal{N} = ((\mathbb{N} \cup \{\infty\}), \min, +)$ are always equivalent to one-way automata, [5]. It is worthwhile observing that the tropical semiring is isomorphic to a sub-semiring of RAT (a^*) in the mapping $n \mapsto a^n a^*$ and that consequently an \mathcal{N} -automaton is a particular case of a two-way transducer whose images have all ultimate period equal to 1. This strengthens the evidence that studying periods is important for two-way transducers.

Second, we exhibit a relation accepted by a 2-way transducer with unary input which cannot be recognized by a sweeping transducer. This result is obtained by counting the images associated to a given word, and showing that no sweeping transducer accepting the relation may produce a sufficient number of images for large inputs.

The paper is organized as follows. Section 2 recalls all the basic notions and results on finite automata and transducers. It is meant to make the paper as self-contained as possible. It also recalls the two Hadamard operations on relations which we introduced in our previous paper. These operations are used in order to algebraically characterize the family of relations accepted by output- and input-unary sweeping transducer (Props. 2.13 and 2.14). Section 3 revisits the well-known properties of the rational subsets of the additive monoid of non-negative integers. The emphasis is on controlling the ultimate periods under the operation of set sum and Kleene star in order to express it as a function of the input. Section 4 applies these results to give a necessary condition for an output-unary two-way transducer to be equivalent to a sweeping transducer, namely, the images associated to an input by such a sweeping transducer should have a period which is bounded by some polynomial in the length of the input (Thm. 4.4). Using this strong property, we exhibit a two-way transducer which is not equivalent to a sweeping transducer (Lem. 4.6), and hence we obtain the separation of the output-unary 2-way and sweeping transducers (Thm. 4.7). In Section 5, we consider the case of input-unary 2-way transducers. A particular relation accepted by a 2-way transducer is exhibited and it is proved that no sweeping transducer may accept it (Lem. 5.2). Hence, we obtain the separation of input-unary 2-way and sweeping transducers (Thm. 5.3). Some concluding remarks are given in Section 6.

2. PRELIMINARIES

2.1. Alphabets, words, languages, rational sets

We assume the reader is familiar with language and automata theory. For the sake of completeness we recall some notions and fix some notations.

The *cardinality* of a set X is denoted $|X|$. An *alphabet* Σ is a non-empty finite set of symbols. The free monoid it generates is denoted by Σ^* , and its elements are *words* over Σ including the *empty word* ϵ . The *length* of a word u is $|u|$. For a symbol $c \in \Sigma$, the *number of occurrences* of c in w is denoted $|w|_c$. The *concatenation* of two words u and v is denoted uv . A *language* is a set of words, *i.e.*, a subset of Σ^* .

An alphabet is *unary* if it is a singleton. A *unary word* (*resp.* *unary language*) is a word (*resp.* language) over a unary alphabet.

Given a monoid $(M, \cdot, 1)$, the family of *rational* subsets denoted $\text{RAT}(M)$ is the least family containing the finite sets and closed under set union ($X \cup Y = \{z \mid z \in X \text{ or } z \in Y\}$), set product ($X \cdot Y = \{xy \mid x \in X, y \in Y\}$) and Kleene star ($X^* = \{x_1 \dots x_p \mid p \in \mathbb{N}, x_i \in X\}$ with the convention $x_1 \dots x_p = 1$ when $p = 0$).

In this paper, we are mainly interested in the monoid $M = \Sigma^* \times \Delta^*$. Its subsets are called *relations*. Given $R \subseteq \Sigma^* \times \Delta^*$ and $u \in \Sigma^*$ we denote by $R(u)$ the *image of u by R* , *i.e.*, the language $\{v \in \Delta^* \mid (u, v) \in R\}$.

2.2. Two-way finite automata

We fix an alphabet Σ , called *input alphabet*, and let \triangleright and \triangleleft be two special symbols which do not belong to Σ , called respectively *left* and *right endmarkers*. The set $\Sigma \cup \{\triangleright, \triangleleft\}$ is denoted by $\overline{\Sigma}$.

Definition 2.1. A *2-way finite automaton* (or simply *automaton* if not otherwise stated) A over Σ is a tuple (Q, q_-, Q_+, δ) , where Q is a finite set of *states*, $q_- \in Q$ is the *initial state*, $Q_+ \subseteq Q$ is the set of *accepting states* and δ is the set of *transitions*, a subset of $Q \times \overline{\Sigma} \times \{-1, 0, 1\} \times Q$, with the restriction that it does not contain any transition of the form $(q, \triangleright, -1, q')$ or $(q, \triangleleft, +1, q')$ for any $q, q' \in Q$.

The *size* of A , denoted $\text{SIZE}(A)$, is its number of states, *i.e.*, $\text{SIZE}(A) = |Q|$.

We recall the dynamics of the device. Given an input word $u = u_1 \dots u_n$ on Σ we augment it to $\tilde{u} = u_0 \cdot u_1 \dots u_n \cdot u_{n+1}$ where $u_0 = \triangleright$ and $u_{n+1} = \triangleleft$. The automaton starts the computation with the word \tilde{u} written

on the tape, the input head positioned on the leftmost cell scanning u_0 , and in state q_- . At each step, the automaton reads the input symbol $a \in \overline{\Sigma}$ scanned by the head, and according to its current state q chooses a direction d and a state q' with $(q, a, d, q') \in \delta$. Then it enters the state q' and moves its head according to d . The automaton *accepts* the input word u if it eventually enters an accepting state at the rightmost position, u_{n+1} . Because of the restrictions on transition set, the input head cannot move out of \tilde{u} . The set of all words accepted by the automaton is the *language accepted*. Two automata are *equivalent* if they accept the same language.

Now we consider some restricted versions of finite automata. An automaton is *1-way* (*resp.* *restless*) if no transition is of the form $(q, c, -1, q')$ (*resp.* $(q, c, 0, q')$) for some $q, q' \in Q$ and $c \in \overline{\Sigma}$. It is *sweeping* if the input head changes direction when scanning an endmarker only. The automaton is *deterministic* if for each pair (q, a) in $Q \times \overline{\Sigma}$ (*resp.* in $Q_+ \times \{\triangleleft\}$), there exists at most one pair (*resp.* no pair) (d, q') in $\{-1, 0, 1\} \times Q$ with $(q, a, d, q') \in \delta$, in other words δ is a (partial) function from $(Q \times \overline{\Sigma}) \setminus (Q_+ \times \{\triangleleft\})$ into $\{-1, 0, 1\} \times Q$. It is well-known that all versions accept the same family, namely the family of regular languages (see, for example, [16, 18]).

2.3. Configurations, runs, traces

The description of the system at a fixed time is given by the current state and the input head position: a *configuration of an automaton A over a word u of length n* is a pair (q, p) where q is a state and p is a *position* of \tilde{u} , *i.e.*, an integer such that $0 \leq p \leq n + 1$. The *initial configuration* is the configuration $(q_-, 0)$. An *accepting configuration* is any configuration $(q, n + 1)$ with $q \in Q_+$. We call *border configuration*, any configuration whose position is equal to 0 or $n + 1$.

From the transition set follows the successor relation on configurations on u . A pair of configurations $((q, p), (q', p'))$ belongs to the *successor relation*, written $(q, p) \rightarrow (q', p')$, if the automaton may enter (q', p') from (q, p) in one step, that is $(q, u_p, (p' - p), q')$ belongs to δ . In particular $(p' - p)$ has to be equal to -1 , 0 or 1 . Observe that the relation depends on the input word. Also, if A is deterministic, then the accepting configurations have no successor.

Definition 2.2. A *run* of A on u is a sequence c_0, c_1, \dots, c_ℓ of successive configurations of A on u , *i.e.*, for each $0 \leq i < \ell$, $c_i \rightarrow c_{i+1}$. If $\ell = 0$ the run is reduced to a single configuration and it is called *trivial*.

A run is *successful* if it starts from the initial configuration and halts in some accepting configuration.

An input word u is accepted by an automaton A if there exists a successful run of A on u .

The following notion is probably superfluous when dealing with automata but it is instrumental when working with transducers.

Definition 2.3. The trace of a run $\mathbf{r} = (q_0, p_0), (q_1, p_1), \dots, (q_\ell, p_\ell)$ of A on u is the sequence $\mathbf{t}_\mathbf{r} = t_1, \dots, t_\ell$ of transitions such that for each $0 < i \leq \ell$, t_i is the witness of $(q_{i-1}, p_{i-1}) \rightarrow (q_i, p_i)$, *i.e.*, $t_i = (q_{i-1}, u_{p_{i-1}}, p_i - p_{i-1}, q_i)$.

As a property of its dynamic, we say that a finite automaton is *unambiguous* if there exists at most one successful run on each input word. Trivially, a deterministic automaton is unambiguous as well.

We now define a particular type of runs. A finite run is a *hit*, if its first and last configurations are both border and if no other configuration is border. Because initial and accepting configurations are border, every successful run is a finite composition of hits.

The *controlled composition of runs* is a partial operator on runs, denoted $@$. Given two runs $\mathbf{r} = c_0, c_1, \dots, c_\ell$ and $\mathbf{r}' = c'_0, c'_1, \dots, c'_{\ell'}$, the controlled composition of \mathbf{r} with \mathbf{r}' exists if and only if $c_\ell = c'_0$ and, in this case, is equal to:

$$\mathbf{r} @ \mathbf{r}' = c_0, c_1, \dots, c_{\ell-1}, c_\ell, c'_1, c'_2, \dots, c'_{\ell'}$$

Remark that the matching configuration at the interface of the two runs, has been collapsed into one. The equivalent notation $\mathbf{r} @ c_\ell @ \mathbf{r}'$ is convenient for explicitly naming the matching configuration.

2.4. Two-way transducers

Here Σ and Δ are two fixed *input* and *output* alphabets. Two-way transducers are two-way finite automata which are provided with the ability to output symbols during the computation. A natural way to define such machines is to add a function that maps every transition into some kind of output. At each step, the machine performs a transition, and produces an output.

Definition 2.4. A 2-way *transducer* (or simply *transducer* if not otherwise stated) is a pair $T = (A, \phi)$ where A is an automaton over Σ with transition set δ and where ϕ is a *production function* which is a mapping of δ into the set of non-empty rational subsets of Δ^* . Its *size* is $\text{SIZE}(T) = \text{SIZE}(A)$.

Let u be a word in Σ^* and let \mathbf{r} be a run on u of trace $t_1 \dots t_\ell$. The word $v \in \Delta^*$ is *produced* by \mathbf{r} if it belongs to the subset $\phi(t_1) \dots \phi(t_\ell)$. We will also use the notation $\Phi_T(\mathbf{r}) = \phi(t_1) \dots \phi(t_\ell)$ or simply $\Phi(\mathbf{r})$ when the transducer T is understood.

A pair $(u, v) \in \Sigma^* \times \Delta^*$ is accepted by the transducer if v is produced by a successful run on u . The *relation accepted* by T is the set of all such (u, v) .

By a slight abuse of language, we say that a production function is *single-valued* if the image of each transition is a singleton. The transducer T is *deterministic* (*resp.* *unambiguous*), if A is deterministic (*resp.* unambiguous) and ϕ is *single-valued*. It is *1-way* or *restless* or *sweeping* if A is. It is well known that the family of relations accepted by 1-way transducers is the family of rational relations, (*e.g.*, [4], Thm. III. 7.1 [9, 17]).

Theorem 2.5. 1-way transducers accept exactly the family of rational relations.

The family of rational relations is strictly smaller than the family of relations accepted by general transducers, even when both input and output alphabets are unary.

Example 2.6. The relation $\text{UMULT} = \{(a^n, a^{kn}) \mid n, k \in \mathbb{N}\}$ is accepted by the 3-state 2-way restless single-valued sweeping transducer $((\{\vec{q}, \overleftarrow{q}, q_+\}, \overleftarrow{q}, \{q_+\}, \delta), \phi)$ where:

$$\delta = \left\{ (\vec{q}, a, +1, \vec{q}), (\overleftarrow{q}, a, -1, \overleftarrow{q}), (q_+, a, +1, q_+) \right. \\ \left. (\vec{q}, \triangleleft, -1, \overleftarrow{q}), (\overleftarrow{q}, \triangleright, +1, \vec{q}), (\overleftarrow{q}, \triangleright, +1, q_+) \right\}$$

and ϕ maps $(\vec{q}, a, +1, \vec{q})$ to $\{a\}$ and all other transitions to $\{\epsilon\}$ (see Fig. 1).

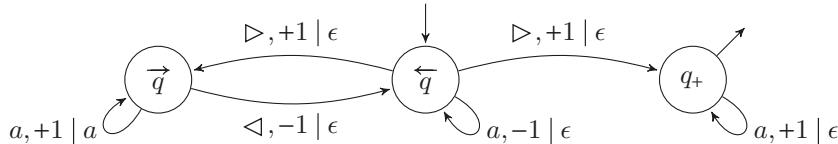


FIGURE 1. A sweeping transducer accepting the relation UMULT (an edge (q, q') is labeled $(s, d \mid w)$ if ϕ maps the transition (q, s, d, q') to w .)

The automaton works in three modes. In state \vec{q} it rewinds the input tape, that is, it moves the input head back to the left endmarker without outputting any symbol. Then, it performs a nondeterministic choice: it either enters state \vec{q} which is used in order to copy the entire input word (observe that the incoming and outgoing transitions are labeled by the left and right endmarkers respectively and that the transition cycling on \vec{q} produce the output $\{a\}$) or it enters state q_+ in order to accept after reaching the right endmarker.

It is easy to show that UMULT is not rational. It follows from Theorem 2.5 that no 1-way transducer can accept it. In ([7], Cor. 1) it was shown that no transducer (A, ϕ) with A being unambiguous, may accept UMULT .

We now show that two-way transducers may be made restless.

Lemma 2.7. Each transducer T admits a computable equivalent restless transducer. Moreover, if T is 1-way (*resp.* sweeping), so is the resulting equivalent restless transducer.

Proof. Let $T = (A, \phi)$ be a transducer of transition set δ . We say that a run is *1-move-at-end* if it is of the form $(q_0, p), (q_1, p), \dots, (q_\ell, p), (q_{\ell+1}, p + d)$ for $\ell \geq 0$, for some position p and some direction $d \in \{-1, +1\}$. The main idea of the proof is to simulate every such 1-move-at-end run by a single restless step $(q_0, p), (q_{\ell+1}, p + d)$. The question of the output generated by this single step is treated in the following.

We consider *0-move runs*, *i.e.* runs whose configurations have all the same position component. Because only one input tape cell is visited, the run does not really depend on the input nor the position but on the current scanned symbol only. The set of 0-move runs which start in state q scanning symbol a and end in state q' (still scanning a) is denoted $\mathbf{R}_{q,q',a}$. The language $L_{q,q',a}$ is defined as the union of the production associated to the runs in $\mathbf{R}_{q,q',a}$, *i.e.*,

$$L_{q,q',a} = \bigcup_{\mathbf{r} \in \mathbf{R}_{q,q',a}} \Phi(\mathbf{r})$$

We prove that $L_{q,q',a}$ is rational. Observe that we can easily obtain a one-way transducer $T_{q,q',a}$ from T such that $T_{q,q',a}$ accepts the relation $\{(a, v) \mid v \in L_{q,q',a}\}$. By Theorem 2.5, the relation is rational and hence, the language $L_{q,q',a}$ is rational by projection.

We now define a new set of transitions δ_1 and its associated production function ϕ_1 (see Fig. 2a). A transition $t = (q, a, d, q'')$ belongs to δ_1 if and only if $d \neq 0$ and there exists a state q' such that $\mathbf{R}_{q,q',a}$ is not empty and (q', a, d, q'') belongs to δ . The rational image of t by ϕ_1 is given by:

$$\phi_1(t) = \bigcup_{q'} (L_{q,q',a} \cdot \phi(q', a, d, q''))$$

Since any configuration (q, p) is a run in $\mathbf{R}_{q,q',a}$, any restless transition t of A belongs to δ_1 and therefore $\phi(t)$ is a subset of $\phi_1(t)$.

Every successful run \mathbf{r} of T can be factorized in

$$\mathbf{r} = \mathbf{r}_0 @ \mathbf{r}_1 @ \dots @ \mathbf{r}_{k-1} @ \mathbf{r}_k$$

where \mathbf{r}_i is a 1-move-at-end run for each $0 \leq i < k$ and \mathbf{r}_k is a 0-move run occurring at the rightmost position (scanning the right endmarker). Using δ_1 and ϕ_1 , we may simulate $\mathbf{r}_0 @ \dots @ \mathbf{r}_{k-1}$ by a restless run \mathbf{r}' . The last factor \mathbf{r}_k is problematic since it may be a non-trivial 0-move run which is not followed by a restless step. We thus need to produce the output in $L_{q,q',\triangleleft}$ for any accepting state q' in the last restless step, *i.e.*, the last step simulating \mathbf{r}_{k-1} . This requires a nondeterministic choice, since the automaton has to guess the presence of the right endmarker one cell to the right.

To this aim, we create a new state q_\triangleleft , which is halting. We then define a transition set δ_\triangleleft and a new production function ϕ_\triangleleft (see Fig. 2b). A transition $t = (q, a, d, q_\triangleleft)$ belongs to δ_\triangleleft if and only if $d = +1$ and there exist two states q' and q'' such that $\mathbf{R}_{q',q'',\triangleleft}$ is not empty, q'' is an accepting state of T and (q, a, d, q') belongs to δ_1 . The image of t by ϕ_\triangleleft is given by:

$$\phi_\triangleleft(q, a, d, q_\triangleleft) = \bigcup_{q' \in Q, q'' \in Q_+} (\phi_1(q, a, d, q') \cdot L_{q',q'',\triangleleft})$$

which is rational.

Our restless transducer simulating $T = (A, \phi)$ is defined as $T' = (A', \phi_1 \cup \phi_\triangleleft)$ where $A' = (Q \cup \{q_\triangleleft\}, q_-, \{q_\triangleleft\}, \delta_1 \cup \delta_\triangleleft)$, where Q and q_- denote respectively the state set and the initial state of A . From any run of T' producing a word v , we can find a run of T producing v and reciprocally. Since the transitions of δ_1 and δ_\triangleleft have the same directions as the corresponding transitions (or 1-move-at-end runs) of T , the construction preserves the property of being one-way or sweeping. \square

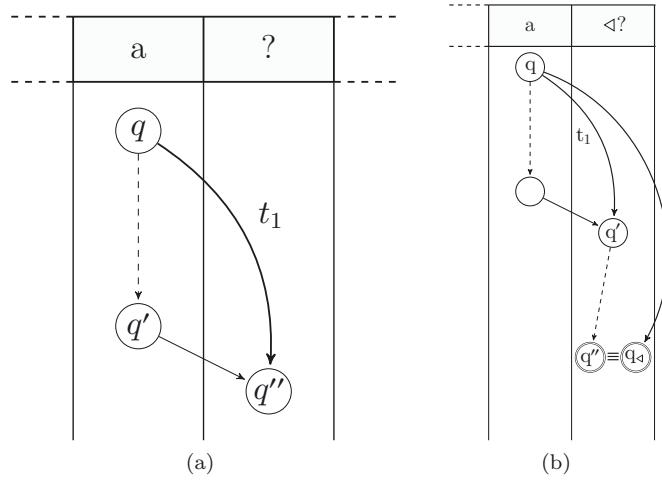


FIGURE 2. Transition in δ_1 (a) and $\delta_<$ (b) for restless simulation of two-way transducers. A dashed arrow holds for a succession of stationary steps and a plain arrow holds for a single step.

2.5. Hadamard relations

The relations accepted by 2-way transducers include strictly the rational relations (see Example 2.6). We recall additional operations that we proved sufficient to express relations accepted by 2-way transducers when both alphabets Σ and Δ are unary [7]. The main result of this paper is that when Δ is unary and Σ arbitrary these operations do not capture the whole family of relations recognized by 2-way transducers but those and only those relations recognized by sweeping transducers. We introduce the *Hadamard operations* and define the family of *Hadamard relations*. All the materials are taken from ([7], Sect. 2), in which relations were represented as formal series. Most of the following results hold under weaker assumptions but the generality is not necessary for our purpose.

Definition 2.8. The *Hadamard operations* of two relations $R, S \subseteq \Sigma^* \times \Delta^*$ are given by:

- the *Hadamard product* (or *H-product*): $R \circledast S$:

$$\forall w \in \Sigma^*, R \circledast S(w) = R(w) \cdot S(w) \quad (\text{set concatenation})$$

- the *Hadamard star* (or *H-star*): $R^{\text{H}*}$:

$$\forall w \in \Sigma^*, R^{\text{H}*}(w) = R(w)^*$$

Under the assumption that Δ is unary the class of rational relations is closed under the Hadamard product, ([17], Thm. III. 3.1).

Theorem 2.9. If Δ is unary then the family $\text{RAT}(\Sigma^* \times \Delta^*)$ is closed under H-product.

However, the H-star of a rational relation is not necessarily rational, even when Σ is unary. Take for example the identity relation on $\Sigma^* \times \Sigma^*$: $\text{ID} = \{(w, w) \mid w \in \Sigma^*\}$. It is trivially rational, but $\text{ID}^{\text{H}*} = \text{uMULT}$ is not rational. Therefore the following defines a broader family.

Definition 2.10. The family of *Hadamard relations*, denoted $\text{HAD}(\Sigma^* \times \Delta^*)$, is the closure of the family $\text{RAT}(\Sigma^* \times \Delta^*)$ by Hadamard operations and union.

Whenever Δ is unary, this family has a simpler characterization.

Proposition 2.11. *Let Σ and Δ be two alphabets with Δ unary. A relation R belongs to $\text{HAD}(\Sigma^* \times \Delta^*)$ if and only if there exist two finite families of rational relations R_i 's and S_i 's, such that: $R = \bigcup_i R_i \textcircled{H} S_i^{\text{H}*}$.*

Proof. Denote by \mathcal{F} the family of relations of the form $\bigcup_i R_i \textcircled{H} S_i^{\text{H}*}$ as in the proposition. By definition, \mathcal{F} is included in $\text{HAD}(\Sigma^* \times \Delta^*)$. Since any rational relation R is equal to $R \textcircled{H} \emptyset^{\text{H}*}$, we have $\text{RAT}(\Sigma^* \times \Delta^*) \subseteq \mathcal{F}$. Thus it suffices to prove that \mathcal{F} is closed under Hadamard operations and union.

The closure under union is trivially obtained from the definition of \mathcal{F} . Let $T = \bigcup_{i \in I} R_i \textcircled{H} S_i^{\text{H}*}$ and $T' = \bigcup_{j \in J} R'_j \textcircled{H} S'_j^{\text{H}*}$ be in \mathcal{F} , and let u be a word in Σ^* . We consider the image of u by the Hadamard product $T \textcircled{H} T'$:

$$\begin{aligned} (T \textcircled{H} T')(u) &= \left(\bigcup_{i \in I} R_i \textcircled{H} S_i^{\text{H}*} \right)(u) \cdot \left(\bigcup_{j \in J} R'_j \textcircled{H} S'_j^{\text{H}*} \right)(u) \\ &= \left(\bigcup_i R_i(u) \cdot S_i(u)^* \right) \cdot \left(\bigcup_j R'_j(u) \cdot S'_j(u)^* \right) \\ &= \bigcup_{i,j} R_i(u) \cdot S_i(u)^* \cdot R'_j(u) \cdot S'_j(u)^* \\ &= \bigcup_{i,j} (R_i(u) \cdot R'_j(u)) \cdot (S_i(u)^* \cdot S'_j(u)^*) && [\text{by commutativity}] \\ &= \bigcup_{i,j} (R_i(u) \cdot R'_j(u)) \cdot (S_i(u) \cup S'_j(u))^* && [\text{by commutativity}] \\ &= \bigcup_{i,j} (R_i \textcircled{H} R'_j)(u) \cdot (S_i \cup S'_j)^{\text{H}*}(u) \\ &= \left(\bigcup_{i,j} (R_i \textcircled{H} R'_j) \textcircled{H} (S_i \cup S'_j)^{\text{H}*} \right)(u) \end{aligned}$$

By Theorem 2.9, each $R_i \textcircled{H} R'_j$ is rational and by definition, each $S_i \cup S'_j$ is also rational. Hence, $T \textcircled{H} T'$ belongs to \mathcal{F} .

We consider now the Hadamard star of T . We claim:

$$T^{\text{H}*} = \left(\bigcup_{i \in I} R_i \textcircled{H} S_i^{\text{H}*} \right)^{\text{H}*} = \bigcup_{X \subseteq I} \left(\bigcup_{i \in X} R_i \textcircled{H} \left(\bigcup_{i \in X} R_i \cup S_i \right) \right)^{\text{H}*} \quad (1)$$

Observe that there are finitely many $X \subseteq I$, and that for each such X , both relations $\textcircled{H}_{i \in X} R_i$ (remember Theorem 2.9) and $\bigcup_{i \in X} R_i \cup S_i$ are rational. This implies that $T^{\text{H}*}$ belongs to \mathcal{F} .

We now prove the equality (1). Let (u, v) be in $T^{\text{H}*}$, that is v belongs to $[(\bigcup_{i \in I} R_i \textcircled{H} S_i^{\text{H}*})(u)]^*$. Thus, $v = v_0 v_1 \dots v_n$ for some $n \in \mathbb{N}$ such that each v_k belongs to $\bigcup_{i \in I} R_i(u) \cdot S_i(u)^*$. For each k we fix an index $i_k \in I$ such that $v_k \in R_{i_k}(u) \cdot S_{i_k}(u)^*$ and we denote by X the set $\{i_k \mid 0 \leq k \leq n\}$. Hence, v belongs to $R_{i_0}(u) \cdot S_{i_0}(u)^* \dots R_{i_n}(u) \cdot S_{i_n}(u)^*$. Using the commutativity of Δ^* , we may reorganize the v_i 's and therefore v belongs to

$$\left(\prod_{i \in X} R_i(u) \right) \cdot \left(\bigcup_{i \in X} R_i(u) \cup S_i(u) \right)^* = \left(\left(\bigcup_{i \in X} R_i \textcircled{H} \left(\bigcup_{i \in X} R_i \cup S_i \right) \right)^{\text{H}*} \right)(u).$$

Reciprocally, let (u, v) belong to $(\textcircled{H}_{i \in X} R_i) \textcircled{H} (\bigcup_{i \in X} R_i \cup S_i)^{\text{H}*}$ for some $X \subseteq I$. For some x and y , we have $v = xy$ where x belongs to $\prod_{i \in X} R_i(u)$ and y belongs to $(\bigcup_{i \in X} R_i(u) \cup S_i(u))^*$. Using commutativity, we may decompose y into $y_0 \dots y_q y_{q+1} \dots y_{q+p}$, where for each $0 \leq h \leq q$ (*resp.* each $q < h \leq q+p$), the word y_h belongs

to $R_{i_h}(u)$ (*resp.* $S_{i_h}(u)$) for some $i_h \in X$. For each $i \in X$ we define w_i as the concatenation of each y_h , with $q < h \leq q + p$, such that $i_h = i$, *i.e.*,

$$w_i = \prod_{q < h \leq q+p \mid i_h=i} y_h$$

In particular, if for no $q < h \leq q + p$ the equality $i_h = i$ holds, then w_i is equal to ϵ . We now decompose the word x into $\prod_{i \in X} x_i$ with each $x_i \in R_i(u)$. Finally, we use commutativity to obtain the equality $v = \prod_{i \in X} (x_i \cdot w_i) \cdot \prod_{0 \leq h \leq q} y_h$. Each $x_i \cdot w_i$ belongs to $(R_i \textcircled{H} S_i^{\text{H}*})(u)$ and each y_h belongs to R_{i_h} and thus to $(R_{i_h} \textcircled{H} S_{i_h}^{\text{H}*})(u)$. Hence, v belongs to $T^{\text{H}*}(u) = (\bigcup_{i \in I} R_i \textcircled{H} S_i^{\text{H}*})^{\text{H}*}(u)$. \square

The Hadamard operations were introduced because they are well-suited to the relations realized by 2-way transducers. Indeed, this family of relations is closed under Hadamard operations.

Proposition 2.12. *If R and S are relations accepted by 2-way transducers, so are the relations $R \textcircled{H} S$ and $R^{\text{H}*}$. Moreover if R and S are accepted by sweeping transducers, then so are $R \textcircled{H} S$ and $R^{\text{H}*}$.*

Proof. We build a 2-way transducer accepting $R \textcircled{H} S$. Our transducer works in three successive modes: (1) it simulates a transducer accepting R ; (2) if the simulation succeed, it rewinds the input tape using a fresh state; (3) it simulates a transducer accepting S and finally accepts if this second simulation succeed.

Now, we build a 2-way transducer accepting $R^{\text{H}*}$. The transducer works as follows. From the initial configuration it performs a nondeterministic choice: it either enters a special state in order to accept at the right endmarker after having crossed the entire input tape; or it starts a direct simulation of a 2-way transducer accepting R . In the second case, if the simulation succeed, it rewinds the input tape using a new state and repeats the previous actions starting from the initial nondeterministic choice. Observe that the automaton of Figure 1 can be seen as those resulting from this construction applied to $R = \text{ID}$. Indeed, as previously observed, $\text{UMULT} = \text{ID}^{\text{H}*}$.

We can easily show that both constructions preserve the property of being sweeping². \square

In particular, thanks to Theorem 2.5, if a relation is Hadamard then it is accepted by a sweeping transducer. The converse happens to be true under the assumption that Δ is unary. Observe that the relations produced by a single hit of a sweeping transducer are rational since during that hit, the computation is 1-way. It follows from ([7], Prop. 4) that every sweeping transducer accepts an Hadamard relation, whenever Δ is unary. Hence, Hadamard relations are characterized by sweeping transducers:

Proposition 2.13. *Let Σ and Δ be two alphabets. If Δ is unary, the family of relations accepted by sweeping transducers over Σ , Δ is equal to $\text{HAD}(\Sigma^* \times \Delta^*)$.*

It can be deduced from ([15], Thm. 15) that the same holds for input-unary sweeping transducers.

Proposition 2.14. *Let Σ and Δ be two alphabets. If Σ is unary, the family of relations accepted by sweeping transducers over Σ , Δ is equal to $\text{HAD}(\Sigma^* \times \Delta^*)$.*

It is shown in Section 4 (see Thms. 4.7 and 5.3 below), that general 2-way transducers accept more than $\text{HAD}(\Sigma^* \times \Delta^*)$ even when Σ or Δ is unary.

² We may also observe that the first construction, that is, those for $R \textcircled{H} S$, also preserves determinism. This is not the case for the second construction.

3. REVISITING THE FAMILY RAT(a^*)

Taking advantage of the observation that (a^*, \cdot, ϵ) is isomorphic with the additive monoid $(\mathbb{N}, +, 0)$ in the mapping $n \mapsto a^n$, we prefer for notational reasons to work in \mathbb{N} . With this identification we may speak of the subset of \mathbb{N} accepted by an automaton over a unary alphabet. From now on instead of working in $\Sigma^* \times a^*$ we will work in the equivalent structure $\Sigma^* \times \mathbb{N}$. All the terminology on the former structure carries over to the latter. Observe that the concatenation in a^* corresponds to the addition in \mathbb{N} . In particular, the set product on \mathbb{N} is denoted $X + Y$. Its neutral element is the singleton $\{0\}$ (or simply denoted 0) and \emptyset is an absorbing element.

First, we introduce some notations. Speaking of the singleton $\{n\}$, we use the abusive but convenient notation n when the context is clear. Hence, for a subset $X \subseteq \mathbb{N}$, we write $n + X$ for $\{n\} + X$. The multiplication of a subset X by a scalar $p \in \mathbb{N}$, *i.e.*, the set $\{px \mid x \in X\}$, is denoted pX . In particular, $p\mathbb{N}$ is the set of all multiples of p . The sum of p copies of X is denoted X^p , for instance $X^3 = X + X + X$. For an integer k , the set of all integers smaller than k is denoted $\llbracket 0, k \rrbracket$. We say that a subset X of \mathbb{N} is *bounded by* k , if $x \in X$ implies $x < k$, *i.e.*, if $X \subseteq \llbracket 0, k \rrbracket$.

3.1. Rational subsets of \mathbb{N}

The following simple result that characterizes regular sets, is a direct consequence of the famous characterization of rational subsets of \mathbb{N} as semilinear sets, *i.e.*, finite unions of linear sets [10].

Proposition 3.1. *A subset X of \mathbb{N} is regular if and only if there exist two integers t and p and two finite sets A and M respectively bounded by t and p such that $X = A \cup (t + M + p\mathbb{N})$.*

If $X = A \cup (t + M + p\mathbb{N})$, for two integers t and p and two subsets A and M respectively bounded by t and p , we say that $A \cup (t + M + p\mathbb{N})$ is a *rat-expression* for X . The integers t and p are respectively *the threshold* and *the period of the rat-expression* or simply *a threshold* and *a period for* X , when the rat-expression is not made precise. It is possible to choose t and p minimal. In this case t and p are called *the threshold* and *the period* of the regular set X . Observe that if $A \cup (t + M + p\mathbb{N})$ is a rat-expression of a finite set X , then $p = 0$ and so $M = \emptyset$; thus $X = A$. Conversely, if X is infinite, then $p > 0$ and $M \neq \emptyset$.

3.2. Equivalent rat-expressions of regular sets

The same regular subset is definable by different rat-expressions with different thresholds and periods. We show how these parameters can be modified.

Lemma 3.2. *Let X be a regular set and let t and p be the threshold and period of some rat-expression of X . Then, for any $u \geq t$, there exists an effectively constructible rat-expression of X with threshold u and period p .*

Proof. Let $A \cup (t + M + p\mathbb{N})$ be a rat-expression of a regular set X and let u be greater than or equal to t . Since $u \geq t$, for some $k \geq 0$ and $0 \leq s < p$ we have $u = t + s + kp$. Define a subset M' of $\{0, \dots, p-1\}$ as follows:

$$M' = \{i - s \mid i \in M \text{ and } i \geq s\} \cup \{p + i - s \mid i \in M \text{ and } i < s\}$$

and define A' as follows:

$$A' = (t + M + p\{0, \dots, k-1\}) \cup (t + \{i \in M \mid i < s\} + kp)$$

Observe that: $X \cap \llbracket 0, t \rrbracket = A$; $X \cap \llbracket t, u \rrbracket = A'$ and $X \cap (u + \mathbb{N}) = u + M' + p\mathbb{N}$. Thus, $X = (A \cup A') \cup (u + M' + p\mathbb{N})$. \square

Lemma 3.3. *Let X be a regular set, and let t and p be the threshold and period of some rat-expression of X . Then, for any $r > 0$, there exists a computable rat-expression of X with threshold t and period rp .*

Proof. Let $A \cup (t + M + p\mathbb{N})$ be a rat-expression of a regular set X and let r be a positive integer. Simply set M' to be equal to the set $M + p[\![0, r]\!]$. We prove $X = A \cup (t + M' + rp\mathbb{N})$. It suffices to prove $t + M + p\mathbb{N} = t + M' + rp\mathbb{N}$.

Let $x = t + m + kp$ for some $m \in M$ and some $k \geq 0$. Then, the Euclidean division of k by r gives $k = qr + \ell$ with $0 \leq \ell < r$. Hence $x = (t + m + \ell p) + qr p$. By definition of M' , $m + \ell p$ belongs to M' , and thus $x \in t + M' + rp\mathbb{N}$.

Reciprocally, if $x = t + m' + qr p$ for some $m' \in M'$ and $q \geq 0$, then there exist $m \in M$ and $0 \leq \ell < r$ such that $m' = m + \ell p$. Thus, $x = t + m + p(\ell + qr)$ belongs to $t + M + p\mathbb{N}$. This concludes the proof. \square

By combining both Lemmas 3.2 and 3.3, a unique threshold and period can be chosen to work with every regular sets of a finite family:

Corollary 3.4. *Let \mathcal{F} be a finite family of regular sets. For each $X \in \mathcal{F}$, let t_X and p_X denote the threshold and the period of some rat-expression of X . Then, there exists for each $X \in \mathcal{F}$ a rat-expression of X with threshold $\max_{X \in \mathcal{F}}(t_X)$ and period $\text{lcm}_{X \in \mathcal{F}}(p_X)$.*

It is then easy to compute the union or the intersection of regular sets. If $A_X \cup (t + M_X + p\mathbb{N})$ and $A_Y \cup (t + M_Y + p\mathbb{N})$ are two rat-expressions, then their union is equal to $(A_X \cup A_Y) \cup (t + (M_X \cup M_Y) + p\mathbb{N})$, which is a rat-expression with the same threshold and the same period. This following particular case is instrumental to the last part of our proof.

Proposition 3.5. *Given a finite family p_1, p_2, \dots, p_n of distinct prime integers, the least period of the set $\bigcup_{0 < i \leq n} p_i\mathbb{N}$ is equal to $\prod_{0 < i \leq n} p_i$.*

Proof. Let X denote the union over $0 < i \leq n$ of $p_i\mathbb{N}$. By Corollary 3.4 and previous observation, $p = p_1 \times \dots \times p_n$ is a period for X .

The minimal period of X divides p . If it is not equal to p then for some p_i the integer $\widehat{p}_i = \frac{p_1 \times \dots \times p_n}{p_i}$ is a period. Then for large enough k , $k p_i \in X$ implies $k p_i + \widehat{p}_i \in X$, i.e., $k p_i + \widehat{p}_i = r p_j$ for some $1 \leq j \leq n$ and some $r \in \mathbb{N}$. We may suppose that k is prime and greater than p . If $i = j$ then the left handside is divisible by p_i thus \widehat{p}_i is divisible by p_i , a contradiction. Otherwise, p_j divides \widehat{p}_i thus $k p_i$, a contradiction. \square

3.3. The sum of regular subsets of \mathbb{N}

By Kleene Theorem we know that the sum of two regular sets is regular. Here we discuss the value of the threshold and the period of the rat-expression of the sum of two subsets of \mathbb{N} . We start by proving an intermediate result:

Proposition 3.6. *Let t and p be a threshold and a period for a regular set X . Let Y be bounded by some $s \in \mathbb{N}$. Then $X + Y$ admits a rat-expression of threshold $t + s$ and period p .*

Proof. For some A bounded by t and some M bounded by p , we have $X = A \cup (t + M + p\mathbb{N})$. Since $Y = \bigcup_{y \in Y} \{y\}$, we have $X + Y = (A + Y) \cup \bigcup_{y \in Y} (y + t + M + p\mathbb{N})$. We fix $y \in Y$. By assumption $t + y < t + s$. By Lemma 3.2, there exist A_y and M_y respectively bounded by $t + s$ and p such that $y + X = A_y \cup ((t + s) + M_y + p\mathbb{N})$. Finally $X + Y = (\bigcup_{y \in Y} A_y) \cup ((t + s) + (\bigcup_{y \in Y} M_y) + p\mathbb{N})$. \square

Lemma 3.7. *Let t, s and p be three integers, and let J and K be two subsets bounded by p . Then there exist A and M , respectively bounded by $t + s + p$ and p , such that*

$$(t + J + p\mathbb{N}) + (s + K + p\mathbb{N}) = A \cup ((t + s + p) + M + p\mathbb{N})$$

Proof. Observe that $(t + J + p\mathbb{N}) + (s + K + p\mathbb{N})$ is equal to $K + (t + s + J + p\mathbb{N})$. Since K is bounded by p , the result follows directly from Proposition 3.6. \square

We are now able to consider the sum of two general regular sets.

Proposition 3.8. Let $A_X \cup (t_X + M_X + p_X \mathbb{N})$ and $A_Y \cup (t_Y + M_Y + p_Y \mathbb{N})$ be the respective rat-expressions of two regular sets X and Y . Fix $t = \max(t_X, t_Y)$ and $p = \text{lcm}(p_X, p_Y)$. Then the regular set $X + Y$ admits a rat-expression of threshold $(2t + p)$ and period p .

Proof. By Corollary 3.4, we may find A'_X and A'_Y bounded by t and M'_X and M'_Y bounded by p such that: $X = A'_X \cup (t + M'_X + p\mathbb{N})$ and $Y = A'_Y \cup (t + M'_Y + p\mathbb{N})$. By distributivity:

$$X + Y = \bigcup \begin{cases} (A'_X + A'_Y) \\ (A'_X + t + M'_Y + p\mathbb{N}) \\ (A'_Y + t + M'_X + p\mathbb{N}) \\ (2t + M'_X + M'_Y + p\mathbb{N}) \end{cases}$$

We consider each of the four subsets separately:

- (1) the set $A_0 = A'_X + A'_Y$ is a finite set bounded by $2t$ and so by $2t + p$;
- (2) the set $A'_X + t + M'_Y + p\mathbb{N}$ can be rewritten, thanks to Proposition 3.6 and Lemma 3.2, as $A_1 \cup ((2t + p) + M_1 + p\mathbb{N})$ with $A_1 \subseteq [0, 2t + p]$ and $M_1 \subseteq [0, p]$;
- (3) similarly the set $A'_Y + t + M'_X + p\mathbb{N}$ is rewritten as $A_2 \cup ((2t + p) + M_2 + p\mathbb{N})$;
- (4) from Lemma 3.7 follows the existence of the sets A_3 and M_3 , respectively bounded by $2t + p$ and p , such that $2t + M'_X + M'_Y + p\mathbb{N} = A_3 \cup (2t + p + M_3 + p\mathbb{N})$.

We set $A = A_0 \cup A_1 \cup A_2 \cup A_3$ and $M = M_1 \cup M_2 \cup M_3$. We have $X + Y = A \cup ((2t + p) + M + p\mathbb{N})$. \square

3.4. The star of regular subsets of \mathbb{N}

The following lemma gives a characterization of star-generated sets of integers.

Lemma 3.9. Let X be a subset of \mathbb{N} . Then, denoting by r the greatest common divisor of the elements of X , i.e., $r = \gcd(X)$, there exist an integer $t \in \mathbb{N}$ and a finite set $A \subseteq [0, t]$ such that $X^* = r(A \cup (t + \mathbb{N}))$. In particular, X^* is regular and is included in $r\mathbb{N}$.

Proof. Whenever $r = \gcd(X) = 1$, it is known that $X^* = A \cup (t + \mathbb{N})$ for some integer t and some subset A bounded by t , (e.g., [1], Thm. 1.0.1). Observe that $t - 1$ when t is minimal is known as the Frobenius number.

We now extend this result to the general case, where r is arbitrary. We define $X_{/r} = \{x \mid rx \in X\}$. We have $\gcd(X_r) = 1$. Since $X^* = r(X_{/r})^*$ and $\gcd(X_r) = 1$, we have $X^* = r(A \cup t + (\mathbb{N}))$. \square

Considering the Kleene star of a regular set, it happens that both the threshold and the period have no simple expressions. However, we are able to bound the value of the period which is enough for our purpose.

Lemma 3.10. Let $A \cup (t + M + p\mathbb{N})$ be a rat-expression of some non-empty regular set X . Then X^* admits a period less than or equal to $\max(t, p)$.

Proof. Let r denote the greatest common divisor of the elements of X . By Lemma 3.9, $X^* = r(K \cup (\ell + \mathbb{N}))$ and thus r is a period of X^* .

Now we prove an upper bound on r in the two disjoint cases: finite or infinite. If X is finite, then X is equal to A and is thus bounded by t . Since r divides all the elements of X (supposed non-empty), we have $r < t$. Else, if X is infinite, then $p > 0$ and $M \neq \emptyset$. For any $x \in t + M$, r divides both x and $x + p$, and thus r divides p . So $r \leq p$. This concludes the proof. \square

4. SWEEPINGNESS WEAKENS 2-WAY TRANSDUCERS EVEN WITH A UNARY OUTPUT ALPHABET

In [7], we proved that when Σ and Δ are unary, the family of relations in $\Sigma^* \times \Delta^*$ accepted by 2-way transducers is exactly the family of Hadamard relations. The crux of the proof seems to rely on the hypothesis that Δ is unary, since this fact is strongly required by the characterization of the family $\text{HAD}(\Sigma^* \times \Delta^*)$ of Proposition 2.11. In fact, for arbitrary Σ and unary Δ , the construction we developed can be extended to show that the relations accepted by various restricted versions of 2-way transducers are in $\text{HAD}(\Sigma^* \times \Delta^*)$: among others, the transducers that are deterministic, unambiguous, functional, or k -valued. We left open the general case where Σ has at least two elements and Δ is unary. Here, we show that in this case the family of relations realized by 2-way transducers strictly contains the family of Hadamard relations.

4.1. Massaging the productions

In this section we give a kind of normal form for transducers with unary output. Thanks to the identification between unary languages and subsets of \mathbb{N} , we may associate to each production function of such transducers a production function that maps transitions into regular subsets of \mathbb{N} .

We show that transducers over Σ and \mathbb{N} admit a simple form:

Lemma 4.1. *Let T be a transducer with transition set δ and production function $\phi : \delta \rightarrow \text{RAT}(\mathbb{N})$. Then there exists an equivalent transducer T' such that the image of each transition by the production function is of the form $t + p\mathbb{N}$ for some non-negative integers t and p . Moreover, if T is 1-way or restless or both, so is the resulting transducer.*

Proof. We fix the transducer $T = (A, \phi)$. By Proposition 3.1, for each transition e of A the language $\phi(e)$ admits a rat-expression $A_e \cup (t_e + M_e + p_e\mathbb{N})$. By decomposing A_e and M_e as finite union of singletons, we obtain:

$$\phi(e) = \left(\bigcup_{a \in A_e} a + 0\mathbb{N} \right) \cup \left(\bigcup_{m \in M_e} (t_e + m) + p_e\mathbb{N} \right)$$

Hence, by indexing the disjoint union $A_e \cup M_e$ by $I_e = \{0, \dots, |A_e| + |M_e| - 1\}$, the set $\phi(e)$ may be written as $\bigcup_{i \in I_e} t_{i,e} + p_{i,e}\mathbb{N}$.

Now we modify the transducer (A, ϕ) into (A', ϕ') in such a way that the transitions distinguish the indices i chosen in I_e . This can easily be done by recording in the finite control of A' which choice has been done at the last transition. Formally, a state of A' is a pair (q, i) where q is a state of A and i is an index in $\bigcup_e I_e$. For each transition $f = (q, a, d, q')$ of A and each index $i \in \bigcup_e I_e$ there are $|I_f|$ transitions: $((q, i), a, d, (q', j))$ for $j \in I_f$. Finally, the image of a transition $((q, i), a, d, (q', j))$ by ϕ' is defined as $t_{j,f} + p_{j,f}\mathbb{N}$. By construction the resulting transducer is equivalent to T . Observe that the directions are kept. \square

4.2. Images of 1-way transducers

Let $R \subseteq \Sigma^* \times \Delta^*$ be a rational relation, *i.e.*, a relation realized by a 1-way transducer. For all words $u \in \Sigma^*$ the set $R(u) = \{v \in \Delta^* \mid (u, v) \in R\}$ is a rational subset of Δ^* , ([17], Thm. IV.1.3). Here, we show that when Δ is unary the collection of all possible images satisfies a uniform property. We keep identifying Δ^* with \mathbb{N} .

Theorem 4.2. *Let Σ be an arbitrary alphabet. Let R be a rational relation in $\Sigma^* \times \mathbb{N}$. Then, there exist two integers t and p such that, for all $w \in \Sigma^*$, the regular language $R(w)$ admits a rat-expression of threshold $t(|w| + 1)$ and period p .*

Proof. By Theorem 2.5, R is accepted by a 1-way transducer $T = (A, \phi)$ as in Definition 2.4, which we can suppose restless by Lemma 2.7. Let w be an input word in Σ^* and let n denote its length. Let \mathcal{R} be the set of all successful runs of T on w . Observe that since T is 1-way restless, every run \mathbf{r} in \mathcal{R} has length $n+2$ (there is exactly one configuration per position, including endmarkers). Thus \mathcal{R} is finite. The image of w is:

$$R(w) = \bigcup_{\mathbf{r} \in \mathcal{R}} \Phi(\mathbf{r})$$

Via Lemma 4.1 we suppose without loss of generality that for each $e \in \delta$, $\phi(e) = t_e + p_e \mathbb{N}$ for some integers t_e and p_e .

We fix one run $\mathbf{r} \in \mathcal{R}$ of trace \mathbf{t} . For each $e \in \delta$, we denote by r_e the number of occurrences of e in \mathbf{t} . By commutativity (recall that X^p denotes the sum of p copies of X),

$$\Phi(\mathbf{r}) = \sum_{e \in \delta} (t_e r_e + (p_e \mathbb{N})^{r_e}) = \left(\sum_{e \in \delta} t_e r_e \right) + \left(\sum_{e \in \delta} (p_e \mathbb{N})^{r_e} \right)$$

Note that $|\mathbf{t}| = n+1$. Thus, $s_{\mathbf{r}} = \sum_{e \in \delta} t_e r_e$ is an integer less than or equal to $m(n+1)$ where $m = \max_{e \in \delta} (t_e)$. Then, define $C_{\mathbf{r}} = \sum_{e \in \delta} (p_e \mathbb{N})^{r_e}$. Denote by $I_{\mathbf{r}}$ the set of transitions e such that $r_e > 0$. Since for any ℓ we have $\ell \mathbb{N} + \ell \mathbb{N} = \ell \mathbb{N}$, the set $C_{\mathbf{r}}$ is equal to $\sum_{e \in I_{\mathbf{r}}} p_e \mathbb{N}$. Observe that there are finitely many possible $I_{\mathbf{r}}$. By Corollary 3.4, there exist two integers k and p such that for each subset I of transitions, there are two sets A_I and M_I , respectively bounded by k and p , such that $\sum_{e \in I} p_e \mathbb{N} = A_I \cup (k + M_I + p \mathbb{N})$. In particular, $C_{\mathbf{r}} = A_{I_{\mathbf{r}}} \cup (k + M_{I_{\mathbf{r}}} + p \mathbb{N})$. Finally:

$$\begin{aligned} \Phi(\mathbf{r}) &= s_{\mathbf{r}} + (A_{I_{\mathbf{r}}} \cup (k + M_{I_{\mathbf{r}}} + p \mathbb{N})) \\ &= (s_{\mathbf{r}} + A_{I_{\mathbf{r}}}) \cup (s_{\mathbf{r}} + k + M_{I_{\mathbf{r}}} + p \mathbb{N}) \end{aligned}$$

As previously claimed, $s_{\mathbf{r}} < m(n+1)$. We can thus find an integer t , independent on n , such that $k + m(n+1) < t(n+1)$. Then, using Lemma 3.2, we can find $B_{\mathbf{r}}$ bounded by $t(n+1)$ and $M'_{\mathbf{r}}$ bounded by p such that $\Phi(\mathbf{r}) = B_{\mathbf{r}} \cup (t(n+1) + M'_{\mathbf{r}} + p \mathbb{N})$.

Now we consider all successful runs of T on w , i.e., all runs in \mathcal{R} . It follows from our previous study:

$$R(w) = \bigcup_{\mathbf{r} \in \mathcal{R}} B_{\mathbf{r}} \cup (t(n+1) + M'_{\mathbf{r}} + p \mathbb{N})$$

and hence, by commutativity and associativity of the set union operation:

$$R(w) = \left(\bigcup_{\mathbf{r} \in \mathcal{R}} B_{\mathbf{r}} \right) \cup \left(t(n+1) + \left(\bigcup_{\mathbf{r} \in \mathcal{R}} M'_{\mathbf{r}} \right) + p \mathbb{N} \right)$$

Because each $B_{\mathbf{r}}$ and $M'_{\mathbf{r}}$ are respectively bounded by $t(n+1)$ and p , so are their respective unions over \mathcal{R} . \square

4.3. Back to 2-way transducers

From the study of Section 3, we are now able to extend Theorem 4.2 to the relations of the special form $R \mathbin{\textcircled{H}} S^{\text{H}*}$ for some rational relations R and S .

Lemma 4.3. *Let Σ be an arbitrary alphabet. Let R and S be two rational relations in $\Sigma^* \times \mathbb{N}$. The regular set $(R \mathbin{\textcircled{H}} S^{\text{H}*})(w)$ admits a period in $\mathcal{O}(|w|)$.*

Proof. By Theorem 4.2, for Z denoting R or S , there exist two integers t_Z and p_Z , such that for every $w \in \Sigma^*$, there are two finite subsets $A_Z(w), M_Z(w) \subseteq \mathbb{N}$ respectively bounded by $t_Z(|w|+1)$ and p_Z that satisfy:

$$Z(w) = A_Z(w) \cup (t_Z(|w|+1) + M_Z(w) + p_Z \mathbb{N})$$

Consider S^{H^*} . By Lemma 3.10, the set $S^{H^*}(w) = (S(w))^*$ admits a period $q_{S,w}$ less than or equal to $\max(t_S(|w|+1), p_S)$. Applying Lemma 3.3, the integer $p_w = p_R \times q_{S,w}$ is a period for both $R(w)$ and $S(w)^*$ and thus for $R(w) + S(w)^* = (R \oplus H^*)^*(w)$ by Proposition 3.8. Observe that $p_w = p_R \times q_{S,w} \leq p_R \times \max(t_S(|w|+1), p_S)$. This concludes the proof. \square

Finally, we prove our main result:

Theorem 4.4. *Let Σ be an arbitrary alphabet. Let R be an Hadamard relation in $\Sigma^* \times \mathbb{N}$. Then there exists an integer k such that for each input word $w \in \Sigma^*$, the regular set $R(w)$ admits a period in $\mathcal{O}(|w|^k)$.*

Proof. Let R be an Hadamard relation in $\Sigma^* \times \mathbb{N}$. Then, for some finite families of rational relations $(X_i)_{0 \leq i < k}$ and $(Y_i)_{0 \leq i < k}$:

$$R = \bigcup_{0 \leq i < k} X_i \oplus Y_i^{H^*}$$

By Lemma 4.3, for every $0 \leq i < k$ there exists an integer c_i such that for every $w \in \Sigma^*$, the set $(X_i \oplus Y_i^{H^*})(w)$ admits a rat-expression of period $p_i(w)$ with $p_i(w) < c_i(|w|+1)$. By Corollary 3.4 and Lemma 3.3, each $(X_i \oplus Y_i^*)(w)$ admits a rat-expression of period $p(w) = \operatorname{lcm}_{0 \leq i < k} p_i(w)$. Therefore, by union, $p(w)$ is a period for $R(w)$ as well. We conclude by observing that

$$p(w) = \operatorname{lcm}_{0 \leq i < k} p_i(w) \leq \prod_{0 \leq i < k} p_i(w) < \prod_{0 \leq i < k} c_i(|w|+1) = \mathcal{O}(|w|^k). \quad \square$$

4.4. Separating output-unary 2-way and sweeping transducers

Theorem 4.4 allows us to prove that the family $\text{HAD}(\Sigma^* \times a^*)$ is strictly included in the family of relations in $\Sigma^* \times a^*$ accepted by 2-way transducers. We define the relation MULTBLOCK on $\Sigma = \{a, \#\}$ by setting for each input word $w \in \Sigma^*$:

$$\text{MULTBLOCK}(w) = \{kn \mid k, n \in \mathbb{N} \text{ and } w \in \Sigma^* \# a^n \# \Sigma^*\}$$

The relation is accepted by a 2-way transducer:

Proposition 4.5. *The relation MULTBLOCK is accepted by a 2-way transducer.*

Proof. We describe the behavior of a 2-way transducer accepting MULTBLOCK (see Fig. 3). The automaton works in three phases: (1) it scans a prefix of the input until it reaches a nondeterministically chosen symbol $\#$; (2) using $\#$'s as endmarkers, it copies the preceding block of successive a 's an arbitrary number of time including 0 (this phase is similar to the behavior of the automaton described in Example 2.6); (3) after a nondeterministic choice, it scans the remaining suffix of the input and accepts at the right endmarker. It should be clear that the relation accepted is MULTBLOCK . \square

We prove now that MULTBLOCK is not Hadamard:

Lemma 4.6. *The relation MULTBLOCK is not Hadamard.*

Proof. By Theorem 4.4, it suffices to prove by contraposition that there exists an infinite sequence of input words $w_n \in \Sigma^*$ of strictly increasing length such that the minimal period p_n of $\text{MULTBLOCK}(w_n)$ is superpolynomial in the length $|w_n|$.

We consider the sequence of words $w_n = \# a^{r_1} \# \dots \# a^{r_n} \#$ where each r_i denotes the i th prime number. By Proposition 3.5, the minimal period p_n for the language $\text{MULTBLOCK}(w_n)$ is equal to $\prod_{0 < i \leq n} r_i$.

At this point we need an instrumental function whose asymptotic behavior is known. The Landau's function $g(m)$ maps every integer m into the largest order of an element of the symmetric group S_m . Equivalently,

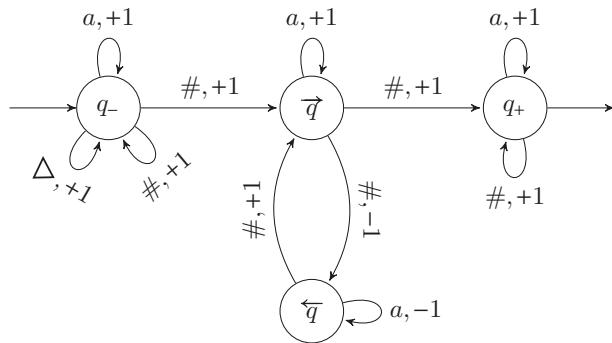


FIGURE 3. The underlying automaton of a 2-way transducer accepting MULTBLOCK. The production function maps the transition $(\overleftarrow{q}, a, -1, \overleftarrow{q})$ to a and all other transitions to ϵ .

it is the largest least common multiple of any partition of m . It is known that $g(m) = e^{(1+o(1))\sqrt{m \ln(m)}}$ [14]. In particular, g is superpolynomial in m .

In our case we have $p_n = g(|w_n|_a) = \prod_{0 < i \leq n} r_i$. Because $|w_n| = |w_n|_a + n + 1$, we obtain that p_n is superpolynomial in $|w_n|$. \square

Observe that for every integer k , the period of the image of $w \in \Sigma^*$ in the restriction $\text{MULTBLOCK} \cap ((\#a^*)^k \# \times \mathbb{N})$ is in $\mathcal{O}(|w|^k)$.

Recall Proposition 2.13 asserting the equivalence between Hadamard relations and relations accepted by sweeping transducers.

Theorem 4.7. *Let Σ and Δ be two alphabets. If Σ has cardinality at least 2 then the family of Hadamard relations in $\text{HAD}(\Sigma^* \times \Delta^*)$, or equivalently the family of relations accepted by sweeping transducers over Σ and Δ , is strictly included in the family of relations accepted by 2-way transducers.*

4.5. A corollary

Recall that the componentwise concatenation of two relations $A_1, A_2 \subseteq \Sigma^* \times \mathbb{N}$ is the relation given by $A_1 \cdot A_2 = \{(u_1 u_2, n_1 + n_2) \mid (u_1, n_1) \in A_1, (u_2, n_2) \in A_2\}$. Define the two relations:

$$\begin{aligned} \text{ERASE} &= \{(w, 0) \mid w \in \Sigma^*\}; \\ \text{MULTONEBLOCK} &= \{(\#a^n \#, kn) \mid n, k \in \mathbb{N} \text{ and } w \in \mathbb{N}\}. \end{aligned}$$

Observe that ERASE is rational, therefore Hadamard, and MULTONEBLOCK is Hadamard (but not rational, compare with the relation UMULT defined in Sect. 2.4). Then we have:

$$\text{MULTBLOCK} = \text{ERASE} \cdot \text{MULTONEBLOCK} \cdot \text{ERASE}$$

The following is a consequence of Lemma 4.6.

Corollary 4.8. *The family of Hadamard relations is not closed under componentwise concatenation, even when the output alphabet is unary.*

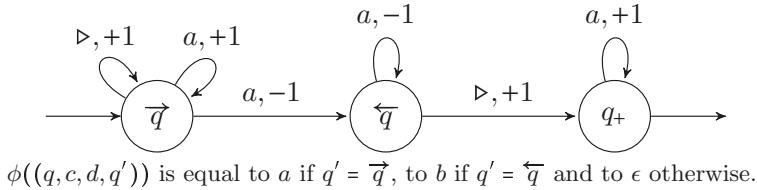


FIGURE 4. A 2-way transducer accepting LR-PREFIX.

5. SWEEPINGNESS WEAKENS 2-WAY TRANSDUCERS EVEN WITH A UNARY INPUT ALPHABET

We know that 2-way transducers are equivalent to sweeping transducers when both the input and the output alphabets are unary. Theorem 4.7 shows that this is not the case anymore, when the output alphabet only is unary. In this section we give an example of a relation in $\{a\}^* \times \{a, b\}^*$ which is accepted by a 2-way transducer but not by a sweeping transducer. This example has also been discussed in [6].

Example 5.1.

$$\text{LR-PREFIX} = \{(a^n, a^p b^p) \mid 0 < p \leq n, n \in \mathbb{N}\}$$

We briefly describe a 2-way transducer accepting LR-PREFIX (see Fig. 4). The device works in three successive phases: (1) it copy a prefix of the input until a nondeterministically chosen point; (2) it scans the prefix backward while outputting a symbol b at each move; (3) it reaches the right endmarker and accepts.

Observe that it perform a nondeterministic choice in state \bar{q} scanning an a . This nondeterminism is strongly required for our purpose, since every deterministic input-unary transducer admits an equivalent sweeping transducer [6] (see Tab. 1). However, observe that the transducer does not admit loops, *i.e.*, no state can be visited twice at the same positions. This kind of transducer is called *simple* in [15] and was studied in [2] in the case of commutative outputs but arbitrary inputs.

The relation LR-PREFIX cannot be accepted by a sweeping transducer.

Lemma 5.2. *No sweeping transducer may accept the relation LR-PREFIX.*

Proof. Suppose there exists a sweeping transducer T accepting LR-PREFIX. By Lemma 2.7, we may assume T is restless. Observe that for each $n > 0$, the language $\text{LR-PREFIX}(a^n)$ has cardinality n , by definition. However, we will prove that on each input, only a bounded number of outputs may be produced by T , a contradiction for large enough inputs.

Let \mathbf{r} be a successful run on some input u . We can decompose \mathbf{r} into three runs

$$\mathbf{r} = \mathbf{r}_1 @ (q, p) @ \mathbf{r}_2 @ (q', p') @ \mathbf{r}_3$$

such that $\phi(\mathbf{r}_1) \in a^*$, and \mathbf{r}_2 is a (q, p) to (q', p') hit (in particular p and p' are border positions, *i.e.*, positions equal to 0 or $|u|+1$) with $\phi(\mathbf{r}_2) \in a^* b^+$ (therefore, $\phi(\mathbf{r}_3) \in b^*$). In other words, we isolate the hit \mathbf{r}_2 during which the first b is output.

Observe that, because T is sweeping, p' is the border position opposite to p . We may thus identify both using the one-bit information $s \in \{\triangleright, \triangleleft\}$, meaning that $u_p = s$. The border opposite to s is denoted \bar{s} and $u_{p'} = \bar{s}$. The tuple $(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3)$ is called *the (q, s, q') -phase decomposition of \mathbf{r}* .

For each states q and q' and each $s \in \{\triangleright, \triangleleft\}$, we define the following relations:

$$\begin{aligned}\text{PREFIX}_{q,s,q'} &= \left\{ (u,v) \mid \begin{array}{l} \text{there exists a successful run on } u \text{ of } (q,s,q')\text{-phase} \\ \text{decomposition } (\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3) \text{ and } v = \phi(\mathbf{r}_1) \end{array} \right\} \\ \text{TRANSIT}_{q,s,q'} &= \left\{ (u,v) \mid \begin{array}{l} \text{there exists a successful run on } u \text{ of } (q,s,q')\text{-phase} \\ \text{decomposition } (\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3) \text{ and } v = \phi(\mathbf{r}_2) \end{array} \right\} \\ \text{SUFFIX}_{q,s,q'} &= \left\{ (u,v) \mid \begin{array}{l} \text{there exists a successful run on } u \text{ of } (q,s,q')\text{-phase} \\ \text{decomposition } (\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3) \text{ and } v = \phi(\mathbf{r}_3) \end{array} \right\}\end{aligned}$$

There are finitely many triples (q, s, q') and by definition:

$$\text{LR-PREFIX} = \bigcup_{q,s,q'} \text{PREFIX}_{q,s,q'} \text{ } \textcircled{+} \text{ } \text{TRANSIT}_{q,s,q'} \text{ } \textcircled{+} \text{ } \text{SUFFIX}_{q,s,q'} \quad (2)$$

Observe that, for each q, q' and s , the three relations defined above have the same domain which is equal to:

$$D_{q,s,q'} = \left\{ u \in a^* \mid \begin{array}{l} \text{there exists a successful run on } u \\ \text{which admits a } (q,s,q')\text{-phase decomposition} \end{array} \right\}$$

It is easy to prove that $D_{q,s,q'}$ is rational. Thus, each relation $\text{TRANSIT}_{q,s,q'}$ is accepted by a 1-way transducer $T_{q,s,q'}$ obtained from T . Denoting p and p' the positions on u such that $u_p = s$ and $u_{p'} = \bar{s}$, the transducer $T_{q,s,q'}$ does not simply simulate the (q,p) to (q',p') hits of T , but should also check that the input belongs to $D_{q,s,q'}$ and that the output contains at least one b . This enforces the simulated (q,p) to (q',p') hit to be the central factor of a (q,s,q') -phase decomposition of some successful run of T .

Observe that, because $|\text{LR-PREFIX}(a^n)| = n$ for any $n > 0$, no successful run of any $T_{q,s,q'}$ on a^n may produce more than n outputs. In particular, every transition used is associated to a finite set of productions. Therefore, the transitions associated with infinite languages are useless and may be removed. Then, we may define k as the maximal length of an output associated with any transition of any transducer $T_{q,s,q'}$, *i.e.*,

$$k = \max \{|y| \in \phi(t) \mid t \text{ transition of } T_{q,s,q'} \text{ with } \phi(t) < \infty \text{ for some } q, s, q'\}.$$

We also fix a constant N greater than the number of states of any $T_{q,s,q'}$ plus one, *i.e.*, $N > \max_{q,s,q'} (\text{SIZE}(T_{q,s,q'})) + 1$.

We fix (q, s, q') . First, we show that $\text{PREFIX}_{q,s,q'}$ and $\text{SUFFIX}_{q,s,q'}$ are functional. Suppose that (u, v) and (u, v') belong to $\text{PREFIX}_{q,s,q'}$. Since the three relations defined above share the same domain, there exist w such that $(u, w) \in \text{TRANSIT}_{q,s,q'} \text{ } \textcircled{+} \text{ } \text{SUFFIX}_{q,s,q'}$, and hence, (u, vw) and $(u, v'w)$ belong to LR-PREFIX . Now, because in each image of LR-PREFIX the number of occurrences of a 's equals those of b 's, we have $|vw|_a = |vw|_b$ and $|v'w|_a = |v'w|_b$. Since both v and v' belong to a^* , we obtain $v = v' = a^{|w|_b - |w|_a}$. Thus, $\text{PREFIX}_{q,s,q'}$ is a function. Similarly, $\text{SUFFIX}_{q,s,q'}$ is a function. It follows that for any $n \in \mathbb{N}$ we have:

$$|(\text{PREFIX}_{q,s,q'} \text{ } \textcircled{+} \text{ } \text{TRANSIT}_{q,s,q'} \text{ } \textcircled{+} \text{ } \text{SUFFIX}_{q,s,q'}) (a^n)| = |\text{TRANSIT}_{q,s,q'}(a^n)|. \quad (3)$$

Moreover, using the same argument, if v and v' in a^*b^+ belong to $\text{TRANSIT}_{q,s,q'}(a^n)$ then we have:

$$|v|_b - |v|_a = |v'|_b - |v'|_a. \quad (4)$$

We will use this property in order to bound by a constant the number of images in each $\text{TRANSIT}_{q,s,q'}(a^n)$. More precisely (recall the two constants k and N defined previously), we prove:

$$\forall (u, v) \in \text{TRANSIT}_{q,s,q'} \quad |v|_a < kN \quad \text{or} \quad |v|_b < kN. \quad (5)$$

Then, because for any u , $\text{TRANSIT}_{q,s,q'}(u) \subseteq a^*b^+$, it follows from (4) that the number of words in $\text{TRANSIT}_{q,s,q'}(u)$ is at most kN . Indeed, for each $0 < i \leq kN$, there exists at most one word $v \in \text{TRANSIT}_{q,s,q'}(u)$ such that $|v|_a = i$ (*resp.* $|v|_b = i$).

We now prove (5). Let n be fixed. A *cycle* is a nontrivial run not visiting the endmarkers³, starting and ending in the same state. Since $\text{TRANSIT}_{q,s,q'}(a^n)$ is included in a^*b^* by definition, no cycle with output in a^*b^* may occur in a successful run of $T_{q,s,q'}$.

Now, we consider two (possibly equal) successful runs \mathbf{r}_1 and \mathbf{r}_2 of $T_{q,s,q'}$ on a^n . Suppose that there are two loops ℓ_1 and ℓ_2 occurring in \mathbf{r}_1 and \mathbf{r}_2 respectively, such that, for some positive integers m_1 and m_2 , we have $a^{m_1} \in \Phi(\ell_1)$ and $b^{m_2} \in \Phi(\ell_2)$. Denote by h_1 and h_2 the length, in terms of head moves, of ℓ_1 and ℓ_2 respectively. Let v_1 and v_2 be two words respectively produced along \mathbf{r}_1 and \mathbf{r}_2 . In particular, by (4), we have $|v_1|_a - |v_1|_b = |v_2|_a - |v_2|_b$. On the pumped input $a^{n+h_1h_2}$, we may find two successful runs: one obtained from \mathbf{r}_1 by repeating h_2 times the cycle ℓ_1 ; the second obtained from \mathbf{r}_2 by repeating h_1 times the cycle ℓ_2 . The words $v'_1 = a^{h_2m_1}v_1$ and $v'_2 = v_2b^{h_1m_2}$ are valid outputs associated to these two runs. Observe that $|v'_1|_a - |v'_1|_b = h_2m_1 + |v_1|_a - |v_1|_b$ and $|v'_2|_a - |v'_2|_b = |v_2|_a - |v_2|_b - h_1m_2$. Hence, using the above constraint on v_1 and v_2 , the two differences are unequal, and therefore, v'_1 and v'_2 violate (4), a contradiction. Thus, for some c equal to a or to b , no cycle occurring in any successful run of $T_{q,s,q'}$ on a^n may output a c .

Fix this c . The maximal number of steps without visiting twice the same state is $\text{SIZE}(T_{q,s,q'}) - 1$. Since steps from and to border configurations do not create cycles, the maximal number of steps in a successful run of $T_{q,s,q'}$ on a^n without cycles is $\text{SIZE}(T_{q,s,q'}) + 1$, which is less than N . Thus, the number of c 's produced along a successful run of $T_{q,s,q'}$ is bounded by kN . This concludes the proof of (5). As explained previously, it follows $|\text{TRANSIT}_{q,s,q'}(a^n)| < kN$.

We conclude by observing that, by the equations (2) and (3), for each n , we have $|\text{LR-PREFIX}(a^n)| < 2kN|Q|^2$ where Q is the state set of T . This is a contradiction, because any input a^n with $n \geq 2kN|Q|^2$ has more associated outputs. \square

It follows that the family of Hadamard relations does not capture the family of relations accepted by 2-way transducers, even when the input alphabet is unary. See the analogy with Theorem 4.7.

Theorem 5.3. *Let Σ and Δ be two alphabets. If Δ has cardinality at least 2 then the family of Hadamard relations in $\text{HAD}(\Sigma^* \times \Delta^*)$, or equivalently the family of relations accepted by sweeping transducer over Σ and Δ , is strictly included in the family of relations accepted by 2-way transducers.*

As for Section 4.5, we deduce a non-closure property of the Hadamard relations.

Corollary 5.4. *The family of Hadamard relations is not closed under componentwise concatenation, even when the input alphabet is unary.*

Proof. Remember that ID denotes the identity relation (see Sect. 2.5), and that ERASE is the relation $\{(w, \epsilon) \mid w \in \Sigma^*\}$ (see Sect. 4.5). We define:

$$\text{RENAME}_{a,b} = \{(a^n, b^n) \mid n \in \mathbb{N}\}$$

Obviously, it is rational. Hence, the Hadamard product: $\text{ID} \circledast \text{RENAME}_{a,b}$ belongs to $\text{HAD}(a^* \times \{a, b\}^*)$. We conclude the proof by observing:

$$\text{LR-PREFIX} = (\text{ID} \circledast \text{RENAME}_{a,b}) \cdot \text{ERASE} \quad \square$$

6. CONCLUSION

Our result proved in [7] claims that on unary input and output alphabets, sweeping transducers have the same recognition power as general 2-way transducers. In this paper, we have shown that the hypothesis of having

³Cycles are used for pumping. Since an endmarker cannot be pumped (it occurs only once on the input tape), we require that no endmarker is visited during the cycles.

both input and output alphabets unary is strongly required. Indeed, we have exhibited two relations, one with a unary output alphabet, the other with a unary input alphabet, that separate the two models (Thms. 4.7 and 5.3). Despite the intuition and the simplicity of the two examples, the proofs involve many intermediate results and some technical material, showing the complexity of the dynamics of 2-way devices. Some of these intermediate results are interesting for their own sake, in particular, the bound on the period of the image of output-unary 1-way (Thm. 4.2) and output-unary sweeping transducers (Thm. 4.4).

REFERENCES

- [1] J. Alfonsín, The Diophantine Frobenius Problem. *Oxford Lecture Series in Mathematics and Its Applications*. OUP, Oxford (2005).
- [2] M. Anselmo, Two-way automata with multiplicity. In Automata, Languages and Programming. Vol. 443 of *Lect. Notes Comput. Sci.* Springer-Verlag, Berlin/Heidelberg (1990) 88–102.
- [3] F. Baschenis, O. Gauwin, A. Muscholl and G. Puppis, One-way definability of sweeping transducer. In *35th IARCS Annual Conference on Foundation of Software Technology and Theoretical Computer Science, FSTTCS (2015) 16–18, 2015, Bangalore, India* (2015) 178–191.
- [4] J. Berstel, *Transductions and context-free languages*. Teubner (1979).
- [5] V. Carnino and S. Lombardy, Tropical Two-Way Automata. In vol. 8705, *Theoretical Computer Science*, Berlin, Heidelberg, Springer Berlin Heidelberg (2014) 195–206.
- [6] C. Choffrut, Sequences of words defined by two-way transducers. *Theoret. Comput. Sci.* **658** (2017) 85–96.
- [7] C. Choffrut and B. Guillon, An Algebraic Characterization of Unary Two-Way Transducers. *Lect. Notes Comput. Sci.* (2014) 196–207.
- [8] R. de Souza, Uniformisation of Two-Way Transducers. In vol. 7810 *Language and Automata Theory and Applications, Lect. Notes Compt. Sci.* Springer Berlin Heidelberg (2013) 547–558.
- [9] S. Eilenberg, *Automata, Languages and Machines*, vol. A. Academic Press (1974).
- [10] S. Eilenberg and M.-P. Schützenberger, Rational sets in commutative monoids. *J. Algebra* **13** (1969) 173–191.
- [11] J. Engelfriet and H.J. Hoogeboom, MSO definable string transductions and two-way finite-state transducers. *ACM Trans. Comput. Logic* **2** (2001) 216–254.
- [12] E. Filiot, O. Gauwin, P.-A. Reynier and F. Servais, From Two-Way to One-Way Finite State Transducers. *CoRR abs/1301.5197* (2013).
- [13] S. Ginsburg and H.G. Rice, Two Families of Languages Related to ALGOL. *J. ACM* **9** (1962) 350–371.
- [14] E. Landau, Über die Maximalordnung der Permutationen gegebenen Grades. *Arch. der Math. u. Phys.* **5** (1903) 92–103.
- [15] S. Lombardy, Weighted two-way automata. In *Seventh Workshop on Non-Classical Models of Automata and Applications – NCMA 2015, Porto, Portugal, August 31 – September 1, (2015). Proceedings* (2015) 37–47.
- [16] M. Rabin and D. Scott, Finite automata and their decision problems. *IBM J. Res. Develop.* **3** (1959) 114–125.
- [17] J. Sakarovitch, *Elements of Automata Theory*. Cambridge University Press (2009).
- [18] J.C. Shepherdson, The reduction of two-way automata to one-way automata. *IBM J. Res. Develop.* **3** (1959) 198–200.

Communicated by M. Holzer.

Received December 19, 2015. Accepted December 16, 2016.