Definability and Regularity in Automatic Structures

Bakhadyr Khoussainov[1], Sasha Rubin[2], and Frank Stephan[3]*

[1] Computer Science Department, The University of Auckland, New Zealand
bmk@cs.auckland.ac.nz
[2] Mathematics Department, The University of Auckland, New Zealand
rubin@math.auckland.ac.nz
[3] National ICT Australia, Sydney Research Laboratory at Kensington, Australia
fstephan@cse.unsw.edu.au

Abstract. An automatic structure $\mathcal{A}$ is one whose domain A and atomic relations are finite automaton (FA) recognisable. A structure isomorphic to $\mathcal{A}$ is called automatically presentable. Suppose R is an FA recognisable relation on A. This paper concerns questions of the following type. For which automatic presentations of $\mathcal{A}$ is (the image of) R also FA recognisable? To this end we say that a relation R is *intrinsically regular* in a structure $\mathcal{A}$ if it is FA recognisable in every automatic presentation of the structure. For example, in every automatic structure all relations definable in first order logic are intrinsically regular. We characterise the intrinsically regular relations of some automatic fragments of arithmetic in the first order logic extended with quantifiers $\exists^\infty$ interpreted as 'there exists infinitely many', and $\exists^{(i)}$ interpreted as 'there exists a multiple of i many'.

1 Introduction

This paper investigates the relationship between regularity, that is FA recognisability, and definability in automatic structures. Roots of this topic go back to the results of Büchi and Elgot in the 1960's who proved the equivalence between regularity and weak monadic second order logic. A recasting of this result says that (the coding of) a relation is regular if and only if the relation it is first order definable in the structure $(\mathbb{N}, +, |_k)$, where $+$ is the addition and $n|_k m$ means that n is a power of k and n divides m. Intimately related is the work of Cobham, Semenov, Muchnik, Bruyére and others that investigates the relationship between regular relations of (coded) natural numbers and definability in certain fragments of arithmetic; see [2] for a good exposition. This paper continues and complements these lines of research by initiating the study of the relationship

* National ICT Australia is funded by the Australian Government's Department of Communications, Information Technology and the Arts and the Australian Research Council through Backing Australia's Ability and the ICT Centre of Excellence Program.

V. Diekert and M. Habib (Eds.): STACS 2004, LNCS 2996, pp. 440–451, 2004.

between regularity and definability in the general setting of arbitrary automatic structures.

Assume one has a structure $\mathcal{A}$ that can be described by means of finite automata. This is formalised in Definition 2.3 that says that there is an encoding of the elements of the structure under which the domain A of the structure and its atomic relations are all regular. Such a structure is called *automatic.* In this case we say that the coded structure is an *automatic presentation* of $\mathcal{A}$. Automatic presentations of $\mathcal{A}$ can be regarded as finite automata implementations of the structure $\mathcal{A}$. For instance, if $k > 1$, then a least-significant-digit-first base k encoding of the natural numbers gives rise to automatic presentations of $(\mathbb{N}, S)$, $(\mathbb{N}, \leq)$, $(\mathbb{N}, +)$ and $(\mathbb{N}, +, |_k)$. Now assume that $R \subset A^m$ is a relation, not necessarily in the language of $\mathcal{A}$. For example, R may be the reachability relation if $\mathcal{A}$ is a graph, or R may be the dependency relation if $\mathcal{A}$ is a group. It may well be the case that in one automatic presentation of $\mathcal{A}$ the relation R is recognised by a finite automaton, and in another automatic presentation it is not. Thus, automata-theoretic properties of the relation R are dependent on the automata that describe $\mathcal{A}$. Our goal is to study those relations in $\mathcal{A}$ that are regular under *all* automatic presentations of $\mathcal{A}$, and to understand which structures ensure a relation is regular in all automatic presentations and which do not. Formally, we introduce the following definition:

Definition 1.1. See [1]. *A relation R is* intrinsically regular *in an automatic structure $\mathcal{A}$ if for every automatic structure $\mathcal{B}$ isomorphic to $\mathcal{A}$ the image of the relation R in $\mathcal{B}$ is regular. Denote by $IR(\mathcal{A})$ the set of intrinsically regular relations in $\mathcal{A}$.*

Thus the intrinsically regular relations in $\mathcal{A}$ are those for which regularity is invariant under all automatic presentations of $\mathcal{A}$. A natural class of intrinsically regular relations is the class of relations definable in the first order logic. We now single out this class of relations in the following definition:

Definition 1.2. *A relation R is* first order (FO) definable *in $\mathcal{A}$ if there exists a first order formula $\phi(\bar{x}, \bar{c})$ with parameters $\bar{c}$ from $\mathcal{A}$ such that $R = \{\bar{a} \mid \mathcal{A} \models \phi(\bar{a}, \bar{c})\}$. Denote by $FO(\mathcal{A})$ the set of all first order definable relations in $\mathcal{A}$.*

A fundamental result of automatic structures is stated as follows.

Fact 1.3. *Let $\mathcal{A}$ be an automatic structure. There exists an algorithm that from a FO definition ϕ in $\mathcal{A}$ of a relation R produces an automaton recognising R. In particular, $FO(\mathcal{A}) \subset IR(\mathcal{A})$.*

A proof may be found in [6] or [3]; in this paper we will use this fact without explicitly referencing it. Naturally, one asks whether the converse holds. It turns out that although this is sufficient for some structures, for instance $(\mathbb{N}, +)$ and $(\mathbb{N}, +, |_m)$, in general it is not.

Extend the FO predicate logic with quantifiers $\exists^{\infty}$ and $\exists^{(i)}$, where $i \in \mathbb{N}$, whose interpretations are as follows. The formula $\exists^{\infty} x\, \phi(x)$ means there are infinitely many x such that $\phi(x)$ holds, and the formula $\exists^{(i)} x\, \phi(x)$ means that there are exactly n elements x such that $\phi(x)$ holds and n is a multiple of i. Denote the logic by $\mathrm{FO}^{\infty,\mathrm{mod}}$. Say that a relation R is *$FO^{\infty,\mathrm{mod}}$ definable* in a structure $\mathcal{A}$ if there is a $\mathrm{FO}^{\infty,\mathrm{mod}}$–formula $\phi(\bar{x}, \bar{a})$, where $\bar{a}$ is a finite tuple of elements, such that $R = \{\bar{c} \mid \mathcal{A} \models \phi(\bar{c}, \bar{a})\}$. Denote by $\mathrm{FO}^{\infty,\mathrm{mod}}(\mathcal{A})$ the set of relations that are $\mathrm{FO}^{\infty,\mathrm{mod}}$ definable in $\mathcal{A}$. Then Fact 1.3 can be extended as follows.

Theorem 3.2. See [3]. [1] *Let $\mathcal{A}$ be an automatic structure. There exists an algorithm that from a $FO^{\infty,\mathrm{mod}}$ definition ϕ of a relation R produces an automaton recognising R. In particular,*

$$\mathrm{FO}^{\infty,\mathrm{mod}}(\mathcal{A}) \subset \mathrm{IR}(\mathcal{A}).$$

Consequently, there is a neat characterisation of the intrinsically regular relations of $(\mathbb{N}, \leq)$ in terms of $\mathrm{FO}^{\infty,\mathrm{mod}}$:

Theorem 3.3.

$$\mathrm{IR}(\mathbb{N}, \leq) = \mathrm{FO}^{\infty,\mathrm{mod}}(\mathbb{N}, \leq) = \mathrm{FO}(\mathbb{N}, \leq, M^2, M^3, \ldots).$$

In order to show that a particular relation is intrinsically regular in a given automatic structure, one needs to provide a mechanism for extracting an automaton recognising the relation from automatic presentations of the structure. A perfect illustration of this is the subset like construction proof of Theorem 3.2. In order to show that a particular relation is not intrinsically regular, one needs to construct automata that, on the one hand, present the structure; and on the other, preclude the existence of automata recognising the given relation. The following theorem shows that the unary relations M^k are not intrinsically regular for the structure $(\mathbb{N}, S)$.

Theorem 4.1. *For every $k \geq 2$, there is an automatic presentation of $(\mathbb{N}, S)$ in which the image of the set M^k is not regular.*

Consequently we have the following partial result.

Corollary 4.2. For $R \subset \mathbb{N}$,

$$R \in \mathrm{IR}(\mathbb{N}, S) \text{ if and only if } R \in \mathrm{FO}^{\infty,\mathrm{mod}}(\mathbb{N}, S).$$

Theorem 4.1 and its proof may be applied to construct automatic structures with pathological properties. The first application is concerned with the reachability problem in automatic graphs. It is known that the reachability problem for automatic graphs is not decidable, see [3]. The underlying reason for this is that such automatic graphs necessarily have infinitely many components. In fact, the reachability problem is decidable if the given graph is automatic and has finitely many components. A natural question is whether or not the reachability relation for automatic graphs with finitely many components can be recognised by finite automata. This is answered in the following corollary:

Corollary 4.3. *There exists an automatic presentation of a graph with exactly two connected components each isomorphic to $(\mathbb{N}, S)$ in which the reachability relation is not regular.*

[1] In a personal communication with the first author, A. Blumensath has mentioned having obtained this result.

The second application of Theorem 4.1 is on the structure $(\mathbb{Z}, S)$, where $\mathbb{Z}$ is the set of all integers and S is the successor function. A *cut* is a set of the form $\{x \in \mathbb{Z} \mid x \geq n\}$, where $n \in \mathbb{Z}$ is fixed. In all previously known automatic presentations of $(\mathbb{Z}, S)$ each cut is a regular set. The corollary below states the existence of a counterexample:

Corollary 4.4. *There exists an automatic presentation of* $(\mathbb{Z}, S)$ *in which no cut is regular.*

Finally, we mention that one of the central topic in modern computable model theory, first initiated by Ash and Nerode in [1], is concerned with understanding the relationship between definability and computability, see [5, Chapter 3] for the current state of the area. For a computable structure $\mathcal{A}$, that is one whose atomic diagram is a computable set, a relation R is *intrinsically recursively enumerable* if in all computable isomorphic copies of $\mathcal{A}$ the relation R is recursively enumerable. In [1] Ash and Nerode show that under some natural conditions put on $\mathcal{A}$, the relation R is intrinsically recursively enumerable if and only if it is definable as an effective disjunction of existential formulas. One may therefore regard the topic of this paper as a refined version of the Ash-Nerode program in which the class of automatic structures is considered rather than the class of all computable structures.

Question 1.4. *Characterise the intrinsically regular relations in* $\mathcal{A}$ *as those definable in a suitable logic of* $\mathcal{A}$.

The results of this paper suggest that the logic is $\mathrm{FO}^{\infty,\mathrm{mod}}$.

The rest of the paper is organised as follows. The next section contains automata preliminaries including the definition of an automatic structure and a description of simple properties of intrinsically regular relations. The remaining sections contain proofs of some of the results stated in the introduction. Due to space constraints some of the proofs are replaced by sketches or completely omitted. The complete proofs can be found in the full version of this paper which is available as a technical report of the Centre for Discrete Mathematics and Theoretical Computer Science in Auckland.

2 Automata Preliminaries

A thorough introduction to automatic structures can be found in [3] and [6]. In this section, familiarity with the basics of finite automata theory is assumed though for completeness and to fix notations, the necessary definitions are included here. A *finite automaton* $\mathcal{A}$ over an alphabet Σ is a tuple (S, ι, Δ, F), where S is a finite set of *states*, $\iota \in S$ is the *initial state*, $\Delta \subset S \times \Sigma \times S$ is the *transition table* and $F \subset S$ is the set of *final states*. A *computation* of $\mathcal{A}$ on a word $\sigma_1\sigma_2 \dots \sigma_n$ ($\sigma_i \in \Sigma$) is a sequence of states, say $q_0, q_1, \dots, q_n$, such that $q_0 = \iota$ and $(q_i, \sigma_{i+1}, q_{i+1}) \in \Delta$ for all $i \in \{0, 1, \dots, n-1\}$. If $q_n \in F$, then the computation is *successful* and we say that automaton $\mathcal{A}$ *accepts* the word. The *language* accepted by the automaton $\mathcal{A}$ is the set of all words accepted by $\mathcal{A}$. In general, $D \subset \Sigma^*$ is *finite automaton recognisable*, or *regular*, if D is the language accepted by a finite automaton $\mathcal{A}$.

Classically finite automata recognise sets of words. The following definitions extends recognisability to relations of arity n, called *synchronous n–tape automata.* Informally a synchronous n–tape automaton can be thought of as a one-way Turing machine with n input tapes. Each tape is regarded as semi-infinite having written on it a word in the alphabet Σ followed by an infinite succession of blanks, $\diamond$ symbols. The automaton starts in the initial state, reads simultaneously the first symbol of each tape, changes state, reads simultaneously the second symbol of each tape, changes state, etc., until it reads a blank on each tape. The automaton then stops and accepts the n–tuple of words if it is in a final state. The set of all n–tuples accepted by the automaton is the relation recognised by the automaton. Here is a formalization. Let $\Sigma_\diamond$ be $\Sigma \cup \{\diamond\}$, where $\diamond \notin \Sigma$.

Definition 2.1. *Write $\Sigma_\diamond$ for $\Sigma \cup \{\diamond\}$ where $\diamond$ is a symbol not in Σ. The* convolution of a tuple *$(w_1, \ldots, w_n) \in \Sigma^{*n}$ is the string $\otimes(w_1, \ldots, w_n)$ of length* $\max_i |w_i|$ *over alphabet $(\Sigma_\diamond)^n$ defined as follows. Its k'th symbol is $(\sigma_1, \ldots, \sigma_n)$ where σ_i is the k'th symbol of w_i if $k \leq |w_i|$ and $\diamond$ otherwise.*

The convolution of a relation *$R \subset \Sigma^{*n}$ is the relation $\otimes R \subset (\Sigma_\diamond)^{n*}$ formed as the set of convolutions of all the tuples in R. That is $\otimes R = \{\otimes w \mid w \in R\}$.*

Definition 2.2. An n–tape automaton *on Σ is a finite automaton over the alphabet $(\Sigma_\diamond)^n$. An n–ary relation $R \subset \Sigma^{*n}$ is* finite automaton recognisable *or* regular *if its convolution $\otimes R$ is recognisable by an n–tape automaton.*

We now relate n–tape automata to structures. A *structure* $\mathcal{A}$ consists of a set A called the *domain* and some relations and operations on A. We may assume that $\mathcal{A}$ only contains relational predicates as the operations can be replaced with their graphs. We write $\mathcal{A} = (A, R_1^A, \ldots, R_k^A, \ldots)$ where R_i^A is an n_i–ary relation on $\mathcal{A}$. The relation R_i are sometimes called basic or atomic relations. We assume that the function $i \rightarrow n_i$ is always a computable one.

Definition 2.3. *A structure $\mathcal{A}$ is* automatic *over Σ if its domain $A \subset \Sigma^*$ is finite automata recognisable, and there is an algorithm that for each i produces a finite automaton recognising the relation $R_i^A \subset \Sigma^{*n_i}$. An isomorphism from a structure $\mathcal{B}$ to an automatic structure $\mathcal{A}$ is an* automatic presentation *of $\mathcal{B}$ in which case $\mathcal{B}$ is called* automatically presentable (*over Σ*). *A structure is called* automatic *if it is automatic over some alphabet.*

Consider the word structure $(\{0,1\}^*, L, R, E, \preceq)$, where for all strings $x, y \in \{0,1\}^*$ we have $L(x) = x0$, $R(x) = x1$, $E(x,y)$ iff $|x| = |y|$, and $\preceq$ is the lexicographical order. It is automatic over Σ. The configuration graphs of Turing machines are examples of automatic structures. Write $\mathbb{N}$ for the set of natural numbers including 0. Examples of automatically presentable structures are $(\mathbb{N}, +)$, $(\mathbb{N}, \leq)$, $(\mathbb{N}, S)$, the group $(\mathbb{Z}, +)$, the order on the rationals $(\mathbb{Q}, \leq)$, and the Boolean algebra of finite or co-finite subsets of $\mathbb{N}$.

Thus an automatic structure is one that is explicitly given by finite automata that recognise the domain and the basic relations of the structure. An automatically presentable structure is one that is isomorphic to some automatic structure. Informally, automatically presentable structures are those that have finite automata implementations. The same structure may have different (indeed infinitely many) automatic presentations. One of our goals is to understand the relationships between different automatic presentations of a given structure and understand how the automata-theoretic properties of relations of this structure change when one varies its automatic presentation. We illustrate the introduced concepts with two examples. The first concerns the standard model of Presburger Arithmetic $(\mathbb{N}, +)$.

Example 2.4. For each $m > 1$ consider the presentation $\mathcal{A}_m$ of $\mathbb{N}$ over the alphabet $\Sigma_m = \{0, \ldots, m-1\}$. Here the natural number $n \in \mathbb{N}$ is represented in A_m as its shortest the least-significant-digit-first base m–representation. The structure $(A_m, +_m)$ is automatic and is isomorphic to $(\mathbb{N}, +)$. Hence, these are automatic presentations of $(\mathbb{N}, +)$. Take any n–ary relation R in $\mathbb{N}$. Assume that R is intrinsically regular. Then the image $R^{(m)}$ of R in $(A_m, +_m)$ is regular. The well-known Cobham-Semenov Theorem, see [2], states that if both $R^{(i)}$ and $R^{(j)}$ are regular for multiplicatively independent i and j, then R is definable in $(\mathbb{N}, +)$. Thus $\mathrm{IR}(\mathbb{N}, +) = \mathrm{FO}(\mathbb{N}, +)$.

The second example concerns an extension of $(\mathbb{N}, +)$.

Example 2.5. Let $|_m$, for $m \geq 2$, be the binary relation where $x|_m y$ if and only if x is a power of m and x divides y. Then the structure $(\mathbb{N}, +, |_m)$ has an automatic presentation $\mathcal{A}_m = (A_m, +_m, D_m)$. So if $R \subset \mathbb{N}^m$ is intrinsically regular for $(\mathbb{N}, +, |_m)$, then its image $R^{(m)}$ is regular in $\mathcal{A}_m$. But a central result of automatic structures is that first order definability in the structure $\mathcal{A}_m$ is equivalent to FA recognisability, see for instance [6]. Hence $R^{(m)}$ is first order definable in $\mathcal{A}_m$, and so R is first order definable in $(\mathbb{N}, +, |_m)$. Thus $\mathrm{IR}(\mathbb{N}, +, |_m) = \mathrm{FO}(\mathbb{N}, +, |_m)$.

3 Intrinsically Regular Relations in $(\mathbb{N}, \leq)$

The linearly ordered set $(\mathbb{N}, \leq)$ has automatic presentations. For example, automatic presentations of $(\mathbb{N}, +)$ are also automatic presentations of $(\mathbb{N}, \leq)$. In this section we study intrinsically regular relations of this structure. Somewhat surprisingly we exhibit intrinsically regular relations of the structure $(\mathbb{N}, \leq)$ that are not definable. We remind the reader that the only first order definable unary relations of $(\mathbb{N}, \leq)$ are finite or co-finite, [4, Theorem 32A].

Let $M^i \subseteq \mathbb{N}$ be the set of all positions in $\mathbb{N}$ that are multiples of i. Then these sets are not definable in $(\mathbb{N}, \leq)$, but are intrinsically regular:

Theorem 3.1. *For every i the unary predicate M^i is intrinsically regular in the structure* $(\mathbb{N}, \leq)$.

Proof. Let $(D, \leq_D)$ be an automatic presentation of $(\mathbb{N}, \leq)$ over Σ. We prove the case when $i = 2$; the case when $i \geq 3$ can be proved in a similar way. Let $E \subseteq D$ be the set of words corresponding to the set of all even natural numbers. Then $x \in E$ iff $\{y \in D \mid y \leq_D x\}$ has odd cardinality. Our goal is to define an automaton over Σ that accepts all such strings x. A rough idea is that the new automaton we want to build calculates the parity of the number of paths in $\mathcal{A}$ with second component fixed at x and accepts x when the parity of the number of successful paths is odd.

Let $\mathcal{A} = (Q_A, \iota_A, \Delta_A, F_A)$ be the automaton over Σ recognising $\leq_D$. We assume that the automaton $\mathcal{A}$ is deterministic. Also, note that since the set $\{y \in D \mid y \leq_D x\}$ is finite for any string $x \in D$, we may assume the following. For each state $s \in Q_A$ there are finitely many strings of the form $(v, \diamond^m)$ that transform the state s into a final state.

Fix a string $x \in D$ and a prefix w of x. For a state $s \in Q_A$ consider all strings v such that $|v| = |w|$ and the automaton $\mathcal{A}$ transforms the string (v, w) to state s from the initial state ι_A. Call these strings (w, s)*–strings.*

The idea in constructing the desired automaton is this. We use the automaton $\mathcal{A}$. Processing the initial prefix w of x, for each state s, we count the parity of the number of (w, s)–strings. We keep a record of only those states s such that the number of (w, s)–strings is odd. By the time we finish processing the string x we have a record of all states s_1, …, s_k such that for each state s_i there are an odd number of (x, s_i)–strings. For each s_i we count the number n_i of strings of the type $\binom{v}{\diamond^m}$, with $m = |v|$, such that the string $\binom{v}{\diamond^m}$ transforms s_i into a final state of $\mathcal{A}$. Then whether or not $x \in E$ can be decided based upon the parity of the numbers n_1, …, n_k. Here is a formal description of the desired automaton $\mathcal{B} = (Q_B, \iota_B, \Delta_B, F_B)$ over $\Sigma_\diamond$:

1. The set Q_B of states of $\mathcal{B}$ are all subsets of Q_A.
2. The initial state ι_B of $\mathcal{B}$ is $\{\iota_A\}$.
3. $\Delta_B(X, \sigma) = Y$, where Y consists of all states $s \in Q_A$ such that there are an odd number of pairs (s', σ') for which $s = \Delta_A(s', \binom{\sigma'}{\sigma})$, $s' \in X$ and $\sigma' \in \Sigma_\diamond$. (Note that Y could be empty).
4. The set of final states F_B is defined as follows. Assume $X = \{s_1, \ldots, s_k\}$ is a subset of Q_A. For each s_i, count the number n_i of all strings of the type $\binom{v}{\diamond^m}$, with $v \in \Sigma^*$, $m = |v|$, such that the string $\binom{v}{\diamond^m}$ transforms s_i into a final state of $\mathcal{A}$. Then $X \in F_B$ if and only if $X \neq \emptyset$ and the number $n_1 + \ldots + n_k$ is odd.

Let $x = \sigma_0 \ldots \sigma_n$ be an input string for $\mathcal{B}$. Let m be the cardinality of the set $\{y \mid y \leq_D x\}$. Let $X_0 = \{\iota_A\}$, X_1, …, X_{n+1} be a run of $\mathcal{B}$ on x. The automaton has the following property $(*)$ that can be proved by induction on $i \geq 1$:

$(*)$ A state s is in X_i if and only if the number of $(\sigma_0 \ldots \sigma_{i-1}, s)$–strings is odd.

Let $X_{n+1} = \{s_1, \ldots, s_n\}$. For each $s_i \in X_{n+1}$ the number of (s_i, x)–strings is odd. Consider the number n_i of all strings of the type $\binom{v}{\diamond^m}$, with $m = |v|$, such that the string $\binom{v}{\diamond^m}$ transforms s_i into a final state of $\mathcal{A}$. From the definition of final states for $\mathcal{B}$, and the inductive assumption on X_n, we see that the cardinality of the set $\{y \mid y \leq_D x\}$ is odd if and only if X_n is non-empty and the number $n_1 + n_2 + \ldots + n_k$ is odd. The theorem is proved. □

As mentioned in the introduction, this proof can be generalised as follows.

Theorem 3.2. *Let $\mathcal{A}$ be an automatic structure. There exists an algorithm that from any $FO^{\infty,\mathrm{mod}}$ definition ϕ of a relation R produces an automaton recognising R. In particular, $FO^{\infty,\mathrm{mod}}(\mathcal{A}) \subset IR(\mathcal{A})$.*

Proof (sketch). Constructing an automata that recognises relation definable by $\exists^{(i)} y\, \psi(\bar{x}, y)$ formula is done in a style similar to the proof of Theorem 3.1. Now note that $\exists^{\infty} y\, \psi(\bar{x}, y)$ is equivalent to $\forall z \exists y (y \preceq z \,\&\, \phi(\bar{x}, y))$, where $\preceq$ is the length-lexicographic ordering on the domain of $\mathcal{A}$. □

Theorem 3.3. *$IR(\mathbb{N}, \leq) = FO^{\infty,\mathrm{mod}}(\mathbb{N}, \leq) = FO(\mathbb{N}, \leq, M^2, M^3, \ldots)$.*

Proof (sketch). By Theorem 3.2, $\mathrm{FO}^{\infty,\mathrm{mod}}(\mathbb{N}, \leq) \subset \mathrm{IR}(\mathbb{N}, \leq)$. So suppose that a relation R is intrinsically regular for $(\mathbb{N}, \leq)$. Then R is regular in every automatic presentation of $(\mathbb{N}, \leq)$. Consider the structure $(1^*, \leq)$ where $\leq$ stands for the ordering induced by the one on the length: $1^n \leq 1^m$ whenever $n \leq m$. This structure is a (unary) automatic presentation of $(\mathbb{N}, \leq)$ and hence the image of R in this presentation is regular. It is shown in [3] that the regular relations over the unary alphabet coincide with those that are first order definable in structure $(\mathbb{N}, \leq, M^2, M^3, \ldots)$. This can be done, for example, via an analysis of finite automata recognising relations over the unary alphabet. Finally, suppose R is first order definable in $(\mathbb{N}, \leq, M^2, M^3, \ldots)$. Since the M^i are $\mathrm{FO}^{\infty,\mathrm{mod}}$ definable in $(\mathbb{N}, \leq)$, then so is R. □

We end the section with a simple application of Theorem 3.2 which gives a generalization of Theorem 3.1. A *tree* $\mathcal{T} = (T, \preceq)$ is a partially ordered set with a least element (the root) and for which every set of the form $\{y \in T \mid y \preceq x\}$ is a finite linear order. The *level* n of $\mathcal{T}$ is the set of all $x \in T$ such that the cardinality of $\{y \in T \mid y \preceq x\}$ is n.

Corollary 3.4. *Let $(T, \preceq)$ be an automatic tree. Given $n \in \mathbb{N}$, the set $\{x \in T \mid x$ is on level $n \cdot m$ for some $m \in \mathbb{N}\}$ is a regular subset of T.*

4 Intrinsic Regularity in $(\mathbb{N}, S)$

Consider the structure $(\mathbb{N}, S)$, where S is the successor function. Our goal is to show that in this structure, all intrinsically regular unary relations are those that are either finite or co–finite. We are also interested in providing automatic presentations of $(\mathbb{N}, S)$ in which some familiar relations are regular and some not. Recall that the finite or co-finite subsets are the only unary relations of this structure that are first order definable, a property that easily follows from elimination of quantifiers [4, Theorem 31G]. The next theorem shows that the set M^k is not intrinsically regular relation in $(\mathbb{N}, S)$, and so by Theorem 3.1 neither is $\leq$.

Theorem 4.1. *For every $k \geq 2$, there is an automatic presentation of $(\mathbb{N}, S)$ in which the image of the set M^k is not regular.*

Proof. Fix $k \geq 2$ and let $\Sigma = \{0, 1, \ldots, k-1\}$. We construct an automatic structure (Σ^*, f) isomorphic to $(\mathbb{N}, S)$. To do this, for any given string $x \in \Sigma^*$, we introduce the following auxiliary notations: $ep(x)$ is the string represented by bits of x at even positions; $op(x)$ is the string represented by bits of x at odd positions; n and m are the lengths of strings $ep(x)$ and $op(x)$, respectively. We may also treat $ep(x)$ and $op(x)$ as natural numbers written in least-significant-digit-first base k, and in particular perform addition on them. For example, if

$x = 0111001$ then $ep(x) = 0101$, $op(x) = 110$, $n = 4$ and $m = 3$; note that $m \leq n \leq m+1$ and $|x| = m+n$. We may regard the string x as the ordered pair of strings, written $\langle ep(x), op(x)\rangle$, and think of $op(x)$ as a parameter. Call strings x for which $ep(x) = k^{n-1}$ *midpoints* and strings for which $ep(x) = 0$ modulo k^n *startpoints.* Now we describe rules defining the function f. In brackets [[like this]] we explain the meaning of each rule if needed. We note in advance that all arithmetic is performed modulo k^n. Define an auxiliary function $\mathrm{next}(x) = ep(x) + kop(x) + k - 1$ modulo k^n.

1. If $n \leq 2$ then $f(x)$ is the successor of x with respect to length-lexicographic ordering.
2. If $\langle \mathrm{next}(x), op(y)\rangle$ is neither a midpoint nor a startpoint then $f(x) = y$, where $ep(y) = \mathrm{next}(x)$ and $op(y) = op(x)$. [[This is the generic case according to which the successor of the string x, regarded as the pair $\langle ep(x), op(x)\rangle$, is $\langle \mathrm{next}(x), op(x)\rangle$.]]
3. If $\langle \mathrm{next}(x), op(y)\rangle$ is a midpoint then $f(x) = y$, where $|y| = |x|$, $ep(y) = ep(x) + \mathrm{next}(\mathrm{next}(x))$ modulo k^n and $op(y) = op(x)$. [[This case says that if adding $\mathrm{next}(x)$ to $ep(x)$ produces a midpoint then the midpoint should be skipped. Note that $ep(y) = ep(x) + 2\mathrm{next}(x)$.]]
4. If $\langle \mathrm{next}(x), op(x)\rangle$ is a startpoint then $f(x) = y$, where $|y| = |x|$, $ep(y) = k^{n-1}$ and $op(y) = op(x)$. [[The successor of the endpoint is the midpoint.]]
5. If $\langle ep(x), op(x)\rangle$ is a midpoint and $op(x) < k^m - 1$ then $f(x) = y$, where $|y| = |x|$, $ep(y) = 0$ and $op(y) = op(x)+1$ modulo k^n. [[This is the case when the parameter $op(x)$ is incremented by 1, and the string $ep(x)$ is initialized to the string consisting of n zeros.]]
6. If $\langle ep(x), op(x)\rangle$ is a midpoint and $op(x) = k^m - 1$ then $f(x) = 0^{n+m+1}$. [[This is the only case when the length of string x increases by one.]]

Now we explain how f acts. Fix $b \in \mathbb{N}$ congruent to $k - 1$ modulo k. For every $a \in \mathbb{N}$ there is a unique number $c \in \{0, 1, \ldots, k^n - 1\}$ such that $a = b \cdot c$ modulo k^n. In other words, every element $c \in \{0, 1, \ldots, k^n - 1\}$ appears exactly once in the sequence $0, b, 2b, 3b, \ldots, (k^n - 1)b$, where elements are taken modulo k^n. Moreover, $k^{n-1}b$ equals k^{n-1} modulo k^n. Hence, $k^{n-1}b$ appears in the middle of this sequence. Let us assume that x is such that $ep(x) = 0$ and let $b = kop(x) + k - 1$. Then by rules 2, 5 and 6, the function f consecutively applied $k^n - 1$ times to $\langle 0, op(x)\rangle$ produces the following sequence:

$$\langle 0, op(x)\rangle, \langle b, op(x)\rangle, \ldots, \langle k^{n-1} - b, op(x)\rangle, \langle k^{n-1} + b, op(x)\rangle,$$
$$\ldots, \langle k^n - b, op(x)\rangle, \langle k^{n-1}, op(x)\rangle.$$

Note that the midpoint $\langle k^{n-1}, op(x)\rangle$ has been removed from the middle of the sequence $\langle 0, op(x)\rangle, \langle b, op(x)\rangle, \ldots, \langle k^n - b, op(x)\rangle$, and placed at the end. Finally rules 3 and 4 imply that f applied to the last string v in the sequence produces the string $\langle 0, op(x) + 1\rangle$ if $op(x) \neq k^m - 1$; otherwise $f(v) = 0^{n+m+1}$. This completes the description of f.

The function f is FA recognisable because all the rules used in the definition of f be can tested by finite automata. It can be checked that (Σ^*, f) is isomorphic to $(\mathbb{N}, S)$, say via mapping $\pi : \Sigma^* \to \mathbb{N}$.

Our goal is to show that the image of the set $M^k = \{x \mid x \text{ is a multiple of } k\}$ is not regular in the described automatic presentation of $(\mathbb{N}, S)$. For this we need to have a finer analysis of the isomorphism π from (Σ^*, f) to $(\mathbb{N}, S)$. Denote by x' the string $\langle 0, op(x)\rangle$. One can inductively check the following for the case that $n \geq 3$.

1. The number $\pi(x')$ is congruent to 0 modulo k for all non-empty strings x.
2. There is a unique $u \leq k^n - 1$ such that $ep(x) = u \cdot (kop(x) + k - 1)$ modulo k^n. Moreover:
   a) If $u < k^{n-1}$ then $\pi(x) = \pi(x') + u$.
   b) If $u > k^{n-1}$ then $\pi(x) = \pi(x') + u - 1$.
   c) If $u = k^{n-1}$ then $\pi(x) = \pi(x') + k^n - 1$.
3. If $ep(y) = 0$ and $op(y) = op(x') + 1 \leq k^m - 1$ then $\pi(y) = \pi(x') + k^n$.

Thus, from the above it is easy to see that x is in the image of M^k iff either $u < k^{n-1}$ and u is congruent to 0 modulo k or $u \geq k^{n-1}$ and u is congruent to 1 modulo k. In order to show that the image of M^k is not regular, consider all the strings x such that n is odd, $ep(x) = 1^n$ (its numerical value is $k^{n+1} - 1$), $op(x) = 0^{m-r}1^r$ (its numerical value is $k^{r+1} - 1$, so that $kop(x) + k - 1 = k^{r+2} - 1$), and $n > r + 4$. Then under these premises for every $r \in \mathbb{N}$ the minimal $n \in \mathbb{N}$ for which $x \in \pi(M^k)$ is when $n = 2r + 5$:

Indeed, $(k^{n-1} + k^{r+2} + 1) \cdot (k^{r+2} - 1) = k^{2r+4} - k^{n-1} - 1$ modulo k^n. So under the assumption that $n = 2r + 5$, this is equal to $-1 = ep(x)$ modulo k^n. Hence $u = k^{n-1} + k^{r+2} + 1 > k^{n-1}$ and so by item 2b above conclude that $\pi(x) = \pi(x') + k^{n-1} + k^{r+1}$ and so $x \in \pi(M^k)$.

For the converse, $(k^{r+2} + 1) \cdot (k^{r+2} - 1) = k^{2r+4} - 1$ modulo k^n. Hence under the assumption that $n < 2r + 5$, this is equal to $-1 = ep(x)$ modulo k^n. Now if further $r + 3 < n - 1$, then $u = k^{r+2} + 1 < k^{n-1}$, and so by item 2a above conclude that $\pi(x) = \pi(x') + k^{r+2} + 1$ and so $x \notin \pi(M^k)$.

Now we can check that $\pi(M^k)$ is not regular. Note that in the presence of $n = 2r + 5$ the assumption that $n > r + 4$ is redundant since $n \leq r + 4$ implies that $r \leq -1$ which contradicts that $r \in \mathbb{N}$. So consider the non regular set

$$Y = \{x \in \Sigma^* \mid ep(x) = 1^n, op(x) = 0^{m-r}1^r, n = 2r + 5\}.$$

It can be defined from $\pi(M^k)$ as the set of all $x \in \Sigma^*$ such that $ep(x) = 1^n$, for some odd n, $op(x) = 0^{m-r}1^r$ for some $m, r \in \mathbb{N}$, $n > r + 4$, $x \in \pi(M^k)$ and if $r + 4 < s < n$ then $(1^s, op(x)) \notin \pi(M^k)$. But since Y is not regular, neither is $\pi(M^k)$, as required. □

Corollary 4.2. *A unary relation $R \subset \mathbb{N}$ is intrinsically regular in $(\mathbb{N}, S)$ if and only if it is in $FO^{\infty,\mathrm{mod}}(\mathbb{N}, S)$.*

Proof. The reverse direction is immediate. For the forward direction it is sufficient to prove that if $R \subset \mathbb{N}$ is intrinsically regular in $(\mathbb{N}, S)$ then it is finite or co-finite; in this case it is in $\mathrm{FO}(\mathbb{N}, S)$ and so certainly in $\mathrm{FO}^{\infty,\mathrm{mod}}(\mathbb{N}, S)$. It can be proved that if R is an eventually periodic set, and if it is infinite and co-infinite, then there is some period p of R such that M_p is first order definable $(\mathbb{N}, S, R)$. Assuming this proceed as follows. Let $R \subset \mathbb{N}$ be intrinsically regular in $(\mathbb{N}, S)$. Since $(1^*, \otimes\{(1^n, 1^{n+1}) \mid n \in \mathbb{N}\})$ is an automatic presentation of $(\mathbb{N}, S)$, R must be eventually periodic. If R is finite or co-finite we are done. Otherwise R is regular in every presentation of $(\mathbb{N}, S)$ and using the fact there exists a period p of R such that M_p is first order definable in $(\mathbb{N}, S, R)$ we get that M_p is also intrinsically regular in $(\mathbb{N}, S)$ contradicting the previous theorem. □

The first application of the results concerns the reachability relation in automatic graphs. The reachability problem for automatic graphs is undecidable, see [3]. The reason for this is that such automatic graphs necessarily have infinitely many components. In fact for automatic graphs with finitely many components the reachability problem is decidable. A natural question is whether or not the reachability relation for automatic graphs with finitely many components can be recognised by finite automata. To answer this question, consider the following graph $\mathcal{G} = (\{0,1\}^*, Edge)$, where $Edge(x, y)$ if and only if $f^2(x) = y$ and f is the function defined in the proof of Theorem 4.1 for $k = 2$. The graph $\mathcal{G}$ is automatic with exactly two infinite components each being isomorphic to $(\mathbb{N}, S)$. One of the components coincides with M^2, and so neither component is regular. Hence, we have the following:

Corollary 4.3. *There exists an automatic presentation of a graph with exactly two connected components each isomorphic to* $(\mathbb{N}, S)$ *for which the reachability relation is not regular.*

A final application of this theorem is on the structure $(\mathbb{Z}, S)$. A *cut* is a set of the form $\{x \in \mathbb{Z} \mid x \geq n\}$, where $n \in \mathbb{Z}$ is fixed.

Corollary 4.4. *There is an automatic presentation of* $(\mathbb{Z}, S)$ *in which no cut is regular.*

Proof (sketch). It is sufficient to find a presentation of $(\mathbb{Z}, S, 0)$ in which $\{x \in \mathbb{Z} \mid x \geq 0\}$ is not regular since every other cut is first order definable from this one. We modify the presentation in the proof of Theorem 4.1 for $k = 2$, by considering the structure $(\{0,1\}^*, g)$, where g is defined using the same notation as before. All arithmetic below is performed modulo 2^n.

1. If $n \leq 2$ then $g(x)$ is the length-lexicographic successor of x.
2. If $(ep(x) + 2op(x) + 1, op(x))$ is neither a midpoint nor a startpoint then $g(x) = y$ with $|x| = |y|$ and $ep(y) = ep(x) + 2op(x) + 1$ and $op(y) = op(x)$.
3. If $(ep(x) + 2op(x) + 1, op(x))$ is a midpoint, then
   a) if $op(x) < 2^m - 1$ then $g(x) = y$ with $|x| = |y|$ and $ep(y) = 0$ and $op(y) = op(x) + 1$.
   b) if $op(x) = 2^m - 1$ then $g(x) = y$ with $|y| = |x| + 1$ and $ep(y) = 0$ and $op(x) = 0$.
4. If $(ep(x) + 2op(x) + 1, op(x))$ is a startpoint, then
   a) if $op(x) < 2^m - 1$ then $g(x) = y$ with $|x| = |y|$ and $ep(y) = 2^{n-1}$ and $op(y) = op(x) + 1$.
   b) if $op(x) = 2^m - 1$ then
      i. if $n = 3$ and $m = 2$ then $g(x) = \epsilon$. Otherwise,

ii. if $n = m + 1$ then $g(x) = y$ with $|ep(y)| = n - 1$, $|op(y)| = m$ and $ep(y) = 2^{n-2}$ and $op(y) = 0$.
iii. if $n = m$ then $g(x) = y$ with $|ep(y)| = n$ and $|op(y)| = m - 1$ and $ep(y) = 2^{n-1}$ and $op(y) = 0$.

Thus, $(\{0,1\}^*, g, \epsilon)$ is an automatic presentation of $(\mathbb{Z}, S, 0)$ in which the cut above 0 is exactly those x such that $u < 2^n - 1$. But if this set were regular then so is the image of M^2 in $(\{0,1\}^*, f)$. □

Finally we mention that there is an automatic presentation of $(\mathbb{N}, S)$ in which $\leq$ is not regular but all the unary relations $M^2, M^3, \ldots$ are regular. This shows that regularity of each of the sets M^i and the successor relation S do not generally imply that the relation $\leq$ is regular. Together with the previous result this theorem says that regularity of $\leq$ is independent of whether or not sets M^i are regular. The proof is available in the full version of this paper.

Theorem 4.5. *The structure* $(\mathbb{N}, S, M^2, M^3, \ldots)$ *has an automatic presentation in which the relation* $\leq$ *is not regular.* □

Acknowledgment. We would like to thank the referees for their suggestions to improve the presentation of the paper and for their thorough reading.

References

1. Christopher J. Ash and Anil Nerode. Intrinsically recursive relations. In *Aspects of effective algebra* (*Clayton, 1979*), pages 26–41. Upside Down A Book Co. Yarra Glen, 1981.
2. Veronique Bruyère, Georges Hansel, Christian Michaux and Roger Villemaire. Logic and p-recognizable sets of integers. *Bull. Belg. Math. Soc.*, 1:191–238, 1994.
3. Achim Blumensath and Erich Grädel. *Automatic Structures*, Proceedings of 15th Symposium on Logic in Computer Science, LICS 2000.
4. Herbert B. Enderton. *A mathematical introduction to logic.* Academic Press, first edition, 1972.
5. Ershov et al. (eds). Handbook of Recursive Mathematics Volume 1. Studies in Logic and the foundations of Mathematics, Elsevier, 1998.
6. Bakhadyr Khoussainov and Anil Nerode. Automatic presentations of structures. *Lecture Notes in Computer Science*, 960:367–392, 1995.