# CDMTCS Research Report Series

# Automatic Linear Orders and Trees (Revised Version)

**Bakhadyr Khoussainov**
Department of Computer Science
University of Auckland
Private Bag 92019, Auckland, New Zealand

**Sasha Rubin**
Department of Mathematics
University of Auckland
Private Bag 92019, Auckland, New Zealand

**Frank Stephan**
National ICT Australia Ltd
Sydney Research Laboratory at Kensington
UNSW Sydney NSW 2052, Australia

CDMTCS-208
November 2003

Centre for Discrete Mathematics and
Theoretical Computer Science

# Automatic Linear Orders and Trees

Bakhadyr Khoussainov, Sasha Rubin and Frank Stephan

November 13, 2003

**Abstract**

We investigate partial orders that are computable, in a precise sense, by finite automata. Our emphasis is on trees and linear orders. We study the relationship between automatic linear orders and trees in terms of rank functions that are related to Cantor-Bendixson rank. We prove that automatic linear orders and automatic trees have finite rank. As an application we provide a procedure for deciding the isomorphism problem for automatic ordinals. We also investigate the complexity and definability of infinite paths in automatic trees. In particular we show that every infinite path in an automatic tree with countably many infinite paths is a regular language.

## 1 Introduction

Consider a class of infinite structures, such as the class of graphs, partial orders, trees, groups, or lattices, etc. A given structure in this class may or may not be computable. If it is one then naturally asks whether or not the structure, or algorithmic problems of the structure, are feasibly computable. In case that 'feasible' means computable by finite automata (see Definition 2.4) one has an *automatic structure.* The automata in this paper operate synchronously on finite words. Using the closure of these automata under boolean operations and projection, one has that the first order theory of an automatic structure is decidable, see for instance [9]. From a computer science point of view this result suggests that automatic structures may be suitable objects that can be effectively queried. The most developed illustration of this is in the related concept of automatic groups in computational group theory [6]. There it is proven that a finitely generated automatic group is finitely presentable and that its word problem is solvable in quadratic time. The general notion of structures presentable by automata has been recently studied in [1, 2, 4, 7, 9, 11]. Throughout this paper we will use the following more general theorem proved in [2] without explicit mention.

**Theorem 1.1** *Given an automatic structure $\mathcal{A}$ and a relation $R$ which is first order definable in $\mathcal{A}$ (with the quantifier $\exists^\infty$ which stands for 'there exist infinitely many'), one can effectively construct an automaton recognising $R$.*

Our work is motivated by the following general problem.

**Problem 1.2** *Given a class of structures $\mathcal{C}$, characterise the isomorphism types of the automatic structures in $\mathcal{C}$.*

The isomorphism type of a structure $\mathcal{A}$ is defined as the set of structures that are isomorphic to $\mathcal{A}$. A satisfactory answer to the problem above would give a non automata theoretic description of those isomorphism types that contain an automatic structure. We note that the isomorphism problem for automatic graphs is $\Sigma_1^1$ complete [10] - and so loosely is as hard as possible. Consequently we restrict the class of structures under consideration. This paper is concerned with the class of partially ordered structures, with an emphasis on trees and linear orderings. A *partial order* (*partial ordering*) is a structure $(A, \preceq)$ such that $\preceq$ is a reflexive, transitive and anti-symmetric binary relation on the domain $A$. A *linear order* $\mathcal{L}$ is a partial order $(L, \leq)$ in which $\leq$ is total, that is $\forall x \forall y (x \leq y \vee y \leq x)$.

Classically linear orderings are characterised in terms of scattered and dense linear orderings as follows. One says that $\mathcal{L}$ is *dense* if for all distinct $a$ and $b$ in $L$ with $a < b$ there exists an $x \in L$ with $a < x < b$. There are only five types of countable dense linear orderings up to isomorphism: the order of rational numbers with or without least and greatest elements, and the order type of the trivial linear order with exactly one element. One says that $\mathcal{L}$ is *scattered* if it does not contain a nontrivial dense subordering. Examples of scattered linear orders are finite sums (Definition 3.1) of cartesian products of $\omega$ (the order type of the natural numbers) and $\mathbb{Z}$ (the order type of the integers). The following theorem is the classical representation of countable linear orderings and can be found in [14, Theorem 4.9].

**Theorem 3.2** *Every countable linear ordering $\mathcal{L}$ can be represented as a dense sum of countable scattered linear orderings.*

The scattered linear orderings can be characterised inductively whereby to each linear order $\mathcal{L}$ one associates a countable ordinal – called the $VD$–rank of $\mathcal{L}$ (Definition 3.3), a version of Cantor-Bendixson rank for topological spaces. One of our results in this paper gives an upper bound on the $FC$–rank (Definition 3.8) of automatic linear orders. For scattered linear orders the $FC$–rank coincides with the $VD$–rank.

**Theorem 4.5** *If $\mathcal{L}$ is an automatic linear order then its $FC$–rank is finite.*

The proof of this theorem generalises a novel technique of Delhomme who gives a full characterisation of automatic ordinals.

**Corollary 4.6** [3] *An ordinal $\alpha$ is automatic if and only if $\alpha < \omega^\omega$.*

Consequently the Cantor-normal-form can be extracted from a presentation of an automatic ordinal.

**Theorem 5.3** *The isomorphism problem for automatic ordinals is decidable.*

A *tree* $\mathcal{T} = (T, \preceq)$ is a partial order that has a minimum element and in which every set of the form $\{y \in T \mid y \preceq x\}$ forms a finite linear order. Elements of trees are called nodes. A node $y \in T$ is an immediate successor of $x \in T$ if $x \prec y$ and there does not exist $z \in T$ for which $x \prec z \prec y$. A tree $\mathcal{T}$ is *finitely branching* if each node $x \in T$ has only finitely many immediate successors. A *path* of a tree $(T, \preceq)$ is a subset $P \subseteq T$ which is linearly ordered, closed downward (that is, whenever $y \in P$ and $x \preceq y$ then $x \in P$)

and maximal (with respect to set theoretic inclusion) with these properties. An *infinite path* is a path $P$ consisting of infinitely many nodes. We are interested in understanding algebraic, model-theoretic as well as computational properties of automatic trees.

We deal with trees by associating to each tree its Kleene-Brouwer ordering. This transformation preserves automaticity, and associates the Cantor-Bendixson rank ($CB$–rank for short) of trees with the $VD$–ranks of the associated linear orders. Informally the $CB$–rank of the tree tells us how big the tree is in terms of ordinals, see for example [8]. This relationship between trees and linear orders gives us the next result.

**Theorem 7.9** *The $CB$–rank of an automatic tree is finite.*

It is known that every infinite finitely branching tree has an infinite path – usually referred to as König's Lemma. The proof of this fact does not produce an infinite path constructively. In fact there are even examples of computable finitely branching trees with *exactly* one infinite path, and that path is *not* computable. Moreover if one omits the assumption that the tree is finitely branching then there are examples of computable trees in which every infinite path is not even arithmetical, see [13]. This negative phenomenon fails dramatically when one considers automatic trees, and not only finitely branching ones.

**Theorem 8.2** *Every infinite automatic finitely branching tree has a regular infinite path.*

We can significantly strengthen this theorem under the assumption that the tree has at most countably many paths. Indeed from Theorem 7.9 we derive that if an automatic finitely branching tree $\mathcal{T}$ has countably many infinite paths then every path of $\mathcal{T}$ is regular (Theorem 8.3). This is because the set of paths in such trees is definable. Moreover one may even omit the assumption that the tree be finitely branching.

**Theorem 8.7** *If an automatic tree has countably many infinite paths then every infinite path in it is regular.*

# 2 Preliminaries

All classical definitions and unproved results on linear orderings can be found in [14]. Countable means finite or countably infinite. All structures are assumed to be countable. Definable means first order definable with the additional quantifier $\exists^\infty$.

A thorough introduction to automatic structures can be found in [1, 9]. A recent survey paper [12] discusses the basic results and possible directions for future work in the area. Familiarity with the basics of finite automata theory is assumed though for completeness and to fix notation the necessary definitions are included here.

A *finite automaton* $\mathcal{A}$ over an alphabet $\Sigma$ is a tuple $(S, \iota, \Delta, F)$, 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*. If a word has a successful computation then 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, $D \subset \Sigma^\star$ is *finite automaton recognisable*, or *regular*, if $D$ is equal to the language accepted by $\mathcal{A}$

for some finite automaton $\mathcal{A}$. An automaton $\mathcal{A}$ is *deterministic* if for every $q \in S$ and $\sigma \in \Sigma$ there is a unique $q' \in S$ such that $(q, \sigma, q') \in \Delta$.

Classically finite automata recognise sets of words. The following definition extends recognisability to relations of arity $n$, by *synchronous $n$–tape automata*. Informally a synchronous $n$–tape automaton can be thought of as a one-way Turing machine with $n$ input tapes [5]. Each tape is regarded as semi-infinite having written on it a word in the alphabet $\Sigma$ 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 formalisation:

**Definition 2.1** Let $\Sigma_\diamond$ be $\Sigma \cup \{\Diamond\}$ where $\Diamond \notin \Sigma$. The *convolution of a tuple* $(w_1, \ldots, w_n) \in (\Sigma^\star)^n$ is the tuple $\otimes(w_1, \ldots, w_n) \in ((\Sigma_\diamond)^n)^\star$ formed by concatenating the least number of blank symbols, $\diamond$, to the right ends of the $w_i$, $1 \leq i \leq n$, so that the resulting words have equal length. The *convolution of a relation* $R \subset (\Sigma^\star)^n$ is the relation $\otimes R \subset ((\Sigma_\diamond)^n)^\star$ formed as the set of convolutions of all the tuples in $R$.

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

For instance let $\leq_p$ be the *prefix* relation. That is for $x, y \in \Sigma^\star$ define $x \leq_p y$ if there exists $z \in \Sigma^\star$ such that $xz = y$. If $z$ is not the empty string $\epsilon$ then $x$ is a *proper prefix* of $y$, written $x <_p y$. So $\leq_p$ is a regular binary relation over $\Sigma$ since for example if $\Sigma = \{0, 1\}$ then $\otimes(\leq_p) = \{\binom{1}{1}, \binom{0}{0}\}^\star \{\binom{\diamond}{1}, \binom{\diamond}{0}\}^\star$.

**Proposition 2.3** [9] *n-tape automata can be effectively determinised and are effectively closed under boolean operations and projection.*

We now relate $n$–tape automata to structures. A *structure* $\mathcal{A}$ consists of a set $A$ called the *domain* and some constants, relations and operations on $A$. We may assume that $\mathcal{A}$ only contains relational predicates as the operations can be replaced with their graphs and constants can be thought of as operations of arity 0. We write $\mathcal{A} = (A, R_1^A, \ldots, R_k^A)$ where $R_i^A$ is an $n_i$–ary relation on $\mathcal{A}$. The *signature* of $\mathcal{A}$ is $(R_1, \ldots, R_k)$. A structure is finite (countably infinite) if its domain has a finite (countably infinite) number of elements. An *isomorphism* between structures $\mathcal{A}$ and $\mathcal{B}$ of the same signature is a bijective mapping $\nu : A \to B$ such that for every relational symbol from the signature, say $R$ of arity $i$, $(a_1, \ldots, a_i) \in R^A$ if and only if $(\nu(a_1), \ldots, \nu(a_i)) \in R^B$ for every tuple $(a_1, \ldots, a_i)$.

**Definition 2.4** A structure $\mathcal{A}$ is *automatic over* $\Sigma$ if its domain $A \subset \Sigma^\star$ and the relations $R_i^A \subset \Sigma^{\star n_i}$ are finite automaton recognisable.

An isomorphism from a structure $\mathcal{B}$ to a structure $\mathcal{A}$ that is automatic over $\Sigma$ is an *automatic presentation* of $\mathcal{B}$ in which case $\mathcal{B}$ is called *automatically presentable* over $\Sigma$. A structure is called *automatic* (*automatically presentable*) if it is automatic (automatically presentable) over some alphabet.

So for instance the structure $(\Sigma^\star, \leq_p)$ is automatic. Examples of automatically presentable structures are Presburger arithmetic $(\mathbb{N}, S, +, 0)$, the group of integers $(\mathbb{Z}, +)$, and the Boolean algebra of finite or co-finite subsets of $\mathbb{N}$.

We now mention some important examples of automatic linear orders. Fix an ordering on $\Sigma$, say $\sigma_1 < \sigma_2 < \dots \sigma_k$. Define $x$ *lexicographically less* than $y$, written $x <_{lex} y$, if either $x$ is a proper prefix of $y$, or else in the first place where they differ the symbol in $x$ is $<$ the symbol in $y$. Then $(\Sigma^\star, <_{lex})$ is an automatic linear order. Also define $x$ *length-lexicographically less* than $y$, written $x <_{llex} y$, if $|x| < |y|$ or else $|x| = |y|$ and $x <_{lex} y$. Then $(\Sigma^\star, <_{llex})$ is an automatic linear order of type $\omega$.

The following fact will be used repeatedly in this paper and is a consequence of the regularity of $<_{lex}$ and $<_{llex}$. Let $\mathcal{A}$ be an automatic structure over $\Sigma$. Then every presentation of $\mathcal{A}$ can be extended to include the regular relations $<_{lex}$ and $<_{llex}$.

Examples of automatically presentable linear orders are $(\mathbb{N}, \leq)$, $(\mathbb{Z}, \leq)$ and the order on rationals $(\mathbb{Q}, \leq)$. Moreover, if $\mathcal{L}_1 = (L_1, \leq_1)$ and $\mathcal{L}_2 = (L_2, \leq_2)$ are automatic linear orders then so are their sum and product. Hence the ordinals $\omega^n$ for every $n \in \mathbb{N}$ are automatically presentable.

Below we present two generic examples of automatic trees.

**Example 2.5** *Let $R$ be a regular language and let $Pref(R)$ be the set of prefixes of strings in $R$. Let $\prec_p$ be the prefix relation. Then the partial orders $(Pref(R), \prec_p)$ and $(R \cup \{\epsilon\}, \prec_p)$ are automatic trees.*

**Example 2.6** *Let $R$ be a regular language. Consider the partial order $\mathcal{T} = (R \cup \{\epsilon\}, \preceq)$, where $x \leq y$ iff $x = y$ or $|x| < |y|$ and $x$ is lexicographically smallest among all $x' \in R$ such that $|x| = |x'|$. Then $\mathcal{T}$ is an automatic tree.*

# 3 Linear order preliminaries

A *partial order* (*partial ordering*) is a pair $(A, \preceq)$ such that $\preceq$ is a reflexive, transitive and anti-symmetric binary relation on the domain $A$. A *linear order* $\mathcal{L}$ is a partial order $(L, \leq)$ in which $\leq$ is total, that is $\forall x \forall y (x \leq y \vee y \leq x)$. If $\mathcal{L}$ is a linear ordering, then unless specified we denote its domain by $L$ and ordering by $\leq_L$ or simply $\leq$. Similarly if $S \subset L$ then we write $\mathcal{S} = (S, \leq_S)$ for the ordering with domain $S$ and ordering $\leq$ restricted to $S$. In this case we say that $\mathcal{S}$ is a subordering of $\mathcal{L}$.

Classically linear orderings are characterised in terms of scattered and dense linear orderings. We say that $\mathcal{L}$ is *dense* if for all distinct $a$ and $b$ in $L$ with $a < b$ there exists an $x \in L$ with $a < x < b$. There are only five types of countable dense linear orderings up to isomorphism: the order of rational numbers with or without least and greatest elements, and the order type of the trivial linear order with exactly one element. We say that $\mathcal{L}$ is *scattered* if it does not contain a nontrivial dense subordering.

Write $\omega$ for the (order) type of the positive integers, $\omega^\star$ for the negative integers, $\zeta$ for the integers, $\eta$ for the rationals and $\mathbf{n}$ for the finite order on $n$ elements. The empty ordering is written $\mathbf{0}$ and the ordering with exactly one element is written $\mathbf{1}$. A subordering $\mathcal{S}$ of $\mathcal{L}$ is an *interval* if for every $x, y \in S$ with $x <_L y$ it is the case that $z \in S$ for every $z \in L$ satisfying $x <_L z <_L y$. An interval is *closed* if it is of the form $\{z \in L \mid x \leq z \leq y\}$ if $x \leq y$ and $\{z \in L \mid y \leq z \leq x\}$ otherwise; either way write $[x, y]$.

**Definition 3.1** Consider a linear order $\mathcal{I}$ as an index set for a set of linear orderings $\{\mathcal{A}_i\}_{i\in I}$. The *$\mathcal{I}$-sum*

$$\mathcal{L} = \Sigma\{\mathcal{A}_i \mid i \in I\}$$

is the linear order with domain $\cup_i A_i$. For $x \in A_i, y \in A_j$ define $x \leq_L y$ if $(i <_I j) \vee (i = j \wedge x \leq_{A_i} y)$.

We refer to the case when $I$ is dense as a *dense sum*. If every $\mathcal{A}_i$ is scattered and $\mathcal{I}$ is scattered then the sum is scattered. If $\mathcal{A}_i = \mathcal{B}$ for every $i \in I$, then the sum is written as a product $\mathcal{B}\mathcal{I}$. For instance $\omega 2$ is $\omega + \omega$. The classical characterisation says that:

**Theorem 3.2** [14, Theorem 4.9] *Every countable linear ordering $\mathcal{L}$ can be represented as a dense sum of countable scattered linear orderings.*

In turn the scattered linear orders can be characterised inductively, where to each linear order one associates an ordinal ranking, called the $VD$–rank. $VD$ stands for very discrete.

**Definition 3.3** For each countable ordinal $\alpha$, define the set $VD_\alpha$ of linear orders inductively as

(1) $VD_0 := \{\mathbf{0}, \mathbf{1}\}$.

(2) $VD_\alpha :=$ all linear orderings formed as $\mathcal{I}$–sums where $\mathcal{I}$ is of the type $\omega, \omega^\star, \zeta$ or $\mathbf{n}$ for some $n < \omega$ and every $\mathcal{A}_i$ is a linear ordering from $\bigcup\{VD_\beta \mid \beta < \alpha\}$.

Define the class $VD$ as the union of the $VD_\alpha$'s. The *$VD$–rank* of a linear ordering $\mathcal{L} \in VD$, written $VD(\mathcal{L})$, is the least ordinal $\alpha$ such that $\mathcal{L} \in VD_\alpha$.

**Example 3.4** *Let $\mathcal{L}_1 = \Sigma\{\zeta + \mathbf{n} \mid n \in \omega\}$, $\mathcal{L}_2 = (\zeta \cdot \zeta) \cdot \zeta$. Then $VD(\mathcal{L}_1) = 2$, $VD(\mathcal{L}_2) = 3$ and $VD(\mathcal{L}_1 + \mathcal{L}_2) = 4$. In general, if $\alpha = \max(VD(\mathcal{L}_1), VD(\mathcal{L}_2))$, then $\alpha \leq VD(\mathcal{L}_1 + \mathcal{L}_2) \leq \alpha + 1$.*

**Example 3.5** *Let $\alpha, \beta$ be countable ordinals. Then $VD(\beta) \leq \alpha$ iff $\beta \leq \omega^\alpha$. In particular, $VD(\omega^\alpha) = \alpha$.*

**Theorem 3.6** [14, Theorem 5.24] *A countable linear ordering $\mathcal{L}$ is scattered if and only if $\mathcal{L}$ is in $VD$.*

There is an alternative definition of ranking that generalises $VD$–rank and includes non-scattered linear orders. We proceed with the definitions.

**Definition 3.7** A *condensation map* is a mapping $c$ from $L$ to non-empty intervals of $L$ such that $c(y) = c(x)$ whenever $y \in c(x)$. The *condensation* of $\mathcal{L}$ is the linear order $c[\mathcal{L}]$ whose domain consists of the collection of non-empty intervals $c(x)$ for $x \in L$ ordered by $c(x) \ll c(y)$ if $\forall x' \in c(x) \forall y' \in c(y)(x' < y')$.

(1) $c^0(x) = \{x\}$ for all $x \in L$.

(2) $c^{\beta+1}(x) = \{y \in L \mid c(c^\beta(x)) = c(c^\beta(y))\}$,

(3) $c^\lambda(x) = \bigcup\{c^\beta(x) \mid \beta < \lambda\}$ for limit ordinal $\lambda$.

Note that in the second item the condensation $c$ is being applied to elements of the form $c^\beta(x)$; in other words $c$ is a mapping with domain $c^\beta[\mathcal{L}]$ to non-empty intervals of $c^\beta[\mathcal{L}]$.

As an illustration we prove that every countable linear ordering can be represented as a dense sum of scattered linear orderings (Theorem 3.2).

**Proof** The mapping $\{y \in L \mid [x, y]$ is scattered$\}$, written $c_S(x)$, is a condensation since if $y \in c_S(x)$ then for all $a$, $[y, a]$ does not contain a dense subordering if and only if $[x, a]$ does not contain a dense subordering. Now $\mathcal{L} = \sum\{a \mid a \in c[\mathcal{L}]\}$. and note that each $a = c_S(x) \in c[\mathcal{L}]$ is scattered. Finally $c_S[\mathcal{L}]$ is dense since for $c_S(x) \ll c_S(y)$, if there is no $z$ with $c_S(x) \ll c_S(z) \ll c_S(y)$ then $[x, y]$ is scattered, contrary to assumption. □

**Definition 3.8** Define $c_{FC}(x)$ as $\{y \in L \mid [x, y]$ is a finite interval of $\mathcal{L}\}$.

Here $FC$ stands for finite condensation and indeed $c_{FC}$ is a condensation. The idea here is that $c^1_{FC}(x)$ is the set of elements of $\mathcal{L}$ that are only finitely far away from $x$; $c^2_{FC}(x)$ is the set of elements of $\mathcal{L}$ that are in intervals of $c_{FC}[\mathcal{L}]$ which themselves are only finitely far away in $c_{FC}[\mathcal{L}]$ from the interval $c^1_{FC}(x)$, etc. The least ordinal $\alpha$ such that $c^\beta_{FC}(x) = c^\alpha_{FC}(x)$ for all $x \in L$ and $\beta \geq \alpha$ is called the *$FC$–rank* of $\mathcal{L}$, written $FC(\mathcal{L})$. *From now on we write $c$ for $c_{FC}$.*

**Example 3.9** *A linear order $\mathcal{L}$ is dense if and only if its $FC$–rank is* 0*. Moreover $\mathcal{L}$ is scattered if and only if $c^\alpha[L] \simeq \mathbf{1}$ for some ordinal $\alpha$.*

The following theorem connects $FC$–ranks and $VD$–ranks of scattered linear orderings.

**Theorem 3.10** [14, Theorem 5.24] *If $\mathcal{L}$ is scattered then its $VD$–rank equals its $FC$–rank.*

If $A \subset L$ then write $c$ to mean that the condensation takes place within the set $L$ and $c_A$ to mean that the condensation takes place relative to $A$. That is, $c_A$ is just $c$ with $\mathcal{A}$ replacing $\mathcal{L}$ in the definition. In this case we write $c_A(x)$ for $x \in A$, $c_A[\mathcal{A}]$ and $\ll_A$. Here are some useful properties that will be used without reference.

**Lemma 3.11** (1) [14, Lemma 5.14] *If $\mathcal{L}$ is scattered and $M \subset L$ then $FC(\mathcal{M}) \leq FC(\mathcal{L})$.*

(2) [14, Lemma 5.13 (2)] *$FC(c^\alpha(x)) \leq \alpha$ and $c^\alpha(x)$ is a scattered interval of $\mathcal{L}$ for every ordinal $\alpha$ and $x \in L$.*

(3) [14, Exercise 5.12 (1)] *If $I$ is an interval of $\mathcal{L}$ then $c^\alpha_I(x) = c^\alpha(x) \cap I$ for every ordinal $\alpha$ and $x \in I$.*

(4) *For every $x, y \in L$, if $[x, y]$ is scattered then $c^\alpha_{[x,y]}(x) = c^\alpha_{[x,y]}(y)$ if and only if $FC([x, y]) \leq \alpha$.*

**Proof** We prove item (4). Let $x, y \in L$ and $\alpha$ be an ordinal. Then by definition $FC([x,y]) \leq \alpha$ means that (†) for every $z \in [x,y]$, $c^{\alpha}_{[x,y]}(z) = c^{\alpha+1}_{[x,y]}(z)$, which necessarily equals $[x,y]$ since $[x,y]$ is scattered. Denote the condition $c^{\alpha}_{[x,y]}(x) = c^{\alpha}_{[x,y]}(y)$ by (††).

Then (†) clearly implies (††) by considering $z \in \{x,y\}$. For the converse suppose (††). We first claim that $c^{\alpha}_{[x,y]}(x) = [x,y]$. Indeed (††) implies that $y \in c^{\alpha}_{[x,y]}(x)$ since $c^{\alpha}_{[x,y]}$ is a condensation, which means that $[x,y]$ is a subset of the interval $c^{\alpha}_{[x,y]}(x)$. But also $c^{\alpha}_{[x,y]}(x) \subset [x,y]$ by item (3). Hence $c^{\alpha}_{[x,y]}(x) = [x,y]$ as claimed. So if $z \in [x,y] = c^{\alpha}_{[x,y]}(x)$ then $c^{\alpha}_{[x,y]}(x) = c^{\alpha}_{[x,y]}(z)$ by the property of being a condensation. Hence $z \in [x,y]$ implies that $c^{\alpha}_{[x,y]}(z) = [x,y]$. In particular then also $c^{\alpha+1}_{[x,y]}(z) = [x,y]$ which implies (†) as required. □

# 4 Ranks of automatic linear orders

We now prove the central technical result, Theorem 4.5, via three propositions that generalise the ideas in [3]. As a matter of convenience we introduce the following variation of rank.

**Definition 4.1** If $\mathcal{L}$ is scattered define its $VD_*$–rank as being the least ordinal $\alpha$ such that $\mathcal{L}$ can be written as a *finite* sum of orderings of $VD$–rank $\leq \alpha$.

For example it is not hard to check that $VD(\omega) = VD_*(\omega) = 1$ and that $\omega 2 + 1$ has $VD$–rank 2 but $VD_*$–rank 1. We list two basic properties.

(1) In general $VD_*(\mathcal{L}) \leq VD(\mathcal{L}) \leq VD_*(\mathcal{L}) + 1$.

(2) $c^{\alpha}[\mathcal{L}]$ is a finite linear order if and only if $VD_*(\mathcal{L}) \leq \alpha$.

**Proposition 4.2** *Let $\mathcal{L} = (L, \leq)$ be a scattered linear ordering. Consider a finite partition of the domain $L = A_1 \cup A_2 \cup \ldots \cup A_k$. Then there exists some $1 \leq i \leq k$ with $VD_*(\mathcal{A}_i) = VD_*(\mathcal{L})$.*

**Proof** It is sufficient to prove the proposition for $k = 2$; the proposition clearly reduces to this case. Thus assume that $A_0 \subset L$ and $A_1 = L \setminus A_0$. We need to show, by induction on $VD_*(\mathcal{L})$, that $VD_*(\mathcal{L}) = VD_*(\mathcal{A}_\epsilon)$ for some $\epsilon \in \{0,1\}$. The case when $VD_*(\mathcal{L}) = 0$ or $VD_*(\mathcal{L}) = 1$ is checked easily. Assume that the proposition is true for all $\mathcal{L}$ such that $VD_*(\mathcal{L}) < \alpha$.

Suppose $VD_*(\mathcal{L}) = \alpha$, that is $\mathcal{L}$ is a finite sum of orders of $VD$–rank at most $\alpha$. In particular at least one of these, call it $\mathcal{M}$, must have $VD$–rank exactly $\alpha$. Then $\mathcal{M}$ is an $I$–sum of linear orders $\{\mathcal{L}_i\}$ of $VD$–rank $< \alpha$, where $I$ is of the type $\omega, \omega^{\star}, \zeta$ or $\mathbf{n}$ for some $n < \omega$. We may assume that $\mathcal{M}$ is chosen so that $I$ is not finite, for if every such $\mathcal{M}$ were a finite sum of orders of $VD$–rank $< \alpha$, then $\mathcal{L}$ would have $VD_*$–rank $< \alpha$. So assume that $I$ is infinite, say of type $\omega$ (the other infinite cases are similar).

For the first case suppose that $\alpha = \beta + 1$. Furthermore we can assume that there are infinitely many $i$ such that $VD_*(\mathcal{L}_i) = \beta$, for otherwise we could write $\mathcal{M}$ as a finite sum of orders of $VD$–rank $\beta$, contrary to assumption. For each such $i$ let $A_{\epsilon,i} = L_i \cap A_\epsilon$, where

$\epsilon \in \{0,1\}$. Applying the induction hypothesis to $\mathcal{L}_i$ we see that there is an $\epsilon \in \{0,1\}$ and infinitely many $j$'s such that $VD_*(\mathcal{A}_{\epsilon,j}) = VD_*(\mathcal{L}_j) = \beta$. Hence $A_\epsilon$ contains a subset which is an $\omega$-sum of linear orders of $VD_*$–rank $\beta$. Therefore, $VD_*(\mathcal{A}_\epsilon) > \beta$, and so $VD_*(\mathcal{A}_\epsilon) = \alpha$ as required.

For the second case, suppose that $\alpha$ is a limit ordinal. So $\mathcal{M}$ is an $\omega$–sum of linear orders $\{\mathcal{L}_i\}$ such the $VD$–rank of each $\mathcal{L}_i$ is less than $\alpha$, and the supremum of the $VD$–ranks of $\mathcal{L}_i$ is $\alpha$. Using the notation of the case above, and applying induction, we see that there is an $\epsilon \in \{0,1\}$ and infinitely may $j$'s such that $VD_*(\mathcal{A}_{\epsilon,j}) = VD_*(\mathcal{L}_j)$, and the supremum of the $VD_*$–ranks of these $\mathcal{A}_{\epsilon,j}$'s is $\alpha$. Then $VD_*(\mathcal{A}_\epsilon) = \alpha$ as required. $\square$

**Proposition 4.3** *Let $\mathcal{L}$ have FC–rank $\alpha$. Then for every $\beta < \alpha$ there exists a closed scattered interval of $\mathcal{L}$ of FC–rank $\beta + 1$.*

**Proof** Fix $\beta < \alpha$. Since $\mathcal{L}$ has $FC$–rank $> \beta$, by definition there is some $x \in L$ such that $c^\beta(x) \neq c^{\beta+1}(x)$. Pick $y \in c^{\beta+1}(x) \setminus c^\beta(x)$. Then $c^\beta(x) \neq c^\beta(y)$ and $c^{\beta+1}(x) = c^{\beta+1}(y)$. Recall that $c^\beta_{[x,y]}$ is the condensation mapping $c^\beta$ within the interval $[x,y]$. Hence $c^\beta_{[x,y]}(x) \neq c^\beta_{[x,y]}(y)$ and $c^{\beta+1}_{[x,y]}(x) = c^{\beta+1}_{[x,y]}(y)$. The first fact implies that $VD([x,y]) > \beta$ and the second fact implies that $VD([x,y]) \leq \beta+1$. Hence $VD([x,y]) = \beta+1$ as required. $\square$

**Proposition 4.4** *The $VD$–rank of every automatic scattered linear ordering is finite.*

**Proof** Suppose $\mathcal{L}$ is automatic scattered linear over $\Sigma^\star$. Let $(Q_\leq, \iota_\leq, \Delta_\leq, F_\leq)$ be a deterministic 2–tape automaton recognising the ordering of $\mathcal{L}$. Similarly let $(Q_A, \iota_A, \Delta_A, F_A)$ be a deterministic 3–tape automaton recognising the definable relation $\{(x,z,y) \mid x \leq z \leq y\}$. We assume the state sets $Q_A$ and $Q_\leq$ are disjoint.

For $x, y \in L$ and $v \in \Sigma^\star$, define $[x,y]_v$ as the set of all $z \in L$ such that $x \leq z \leq y$ and $z$ has prefix $v$. For $|v| \geq |x|, |y|$ define $I(x,v,y) \in Q_A$ and $J(x) \in Q_\leq$ as follows. $I(x,v,y)$ is the state in $Q_A$ reachable from the initial state $\iota_A$ after reading the convolution of $(x,v,y)$, namely $(x\Diamond^n, v, y\Diamond^m)$ where $n, m \geq 0$ are chosen so that the length of each component is exactly $|v|$. That is define $I(x,v,y) := \Delta_A(\iota_A, \otimes(x,v,y))$. Similarly define $J(v) := \Delta_\leq(\iota_\leq, \otimes(v,v))$. Write $K(x,v,y)$ for the ordered pair $(I(x,v,y), J(v))$.

Now if $K(x,v,y) = K(x',v',y')$ then $[x,y]_v$ is isomorphic to $[x',y']_{v'}$ via the map $vw \mapsto v'w$ for $w \in \Sigma^\star$. Indeed the domains are isomorphic since

$$vw \in [x,y]_v$$

if and only if

$$\Delta_A(\Delta_A(\iota_A, \otimes(x,v,y)), \otimes(\epsilon, w, \epsilon)) \cap F_A \neq \emptyset$$

if and only if

$$\Delta_A(\Delta_A(\iota_A, \otimes(x',v',y')), \otimes(\epsilon, w, \epsilon)) \cap F_A \neq \emptyset$$

if and only if

$$v'w \in [x',y']_{v'}.$$

The map preserves the ordering since for $w_1, w_2 \in \Sigma^\star$ such that $vw_1, vw_2 \in [x,y]_v$ and $v'w_1, v'w_2 \in [x',y']_{v'}$ we have

$$vw_1 \leq vw_2$$

if and only if

$$\Delta_{\leq}(\Delta_{\leq}(\iota_{\leq}, \otimes(v,v)), \otimes(w_1,w_2)) \cap F_{\leq} \neq \emptyset$$

if and only if

$$\Delta_{\leq}(\Delta_{\leq}(\iota_{\leq}, \otimes(v',v')), \otimes(w_1,w_2)) \ \cap F_{\leq} \neq \emptyset$$

if and only if

$$v'w_1 \leq v'w_2.$$

Hence then number of isomorphism types of the form $[x,y]_v$ for $|v| \geq |x|, |y|$ is bounded by the number of distinct sets $K(x,v,y)$ which is at most $|Q_A| \times |Q_{\leq}|$, denoted by $d$. In particular (†) there are at most $d$ many $VD_*$–ranks among such intervals of the form $[x,y]_v$.

Now let $[x,y]$ be a closed interval of $\mathcal{L}$. Set $n = \max\{|x|,|y|\}$ and partition $[x,y]$ into the set $[x,y] \cap \Sigma^{<n}$ and the finitely many sets of the form $[x,y]_v$ where $|v| = n$. By Proposition 4.2 one of these intervals has the same $VD_*$–rank as the interval $[x,y]$. Suppose now that the $VD$–rank of $[x,y]$ is at least $2(d+1)$. By Proposition 4.3, for every $1 \leq i \leq 2(d+1)$, $[x,y]$ contains a closed interval of $VD$–rank $i$. So by property (1) of $VD_*$–ranks there are at least $d+1$ many $VD_*$–ranks amongst these intervals; call them $[x_j,y_j]$ for $1 \leq j \leq d+1$. But as was done for $[x,y]$ one has that for every such interval $[x_j,y_j]$, there is a $v_j \in \Sigma^\star$ such that $VD_*([x_j,y_j]) = VD_*([x_j,y_j]_v)$, and so contradicting (†). We conclude that the $VD$–rank of $[x,y]$ is at most $e = 2(d+1)$. So for every $x,y \in L$, $c^e(x) = c^e(y)$ and so $VD(\mathcal{L}) \leq e$ as required. □

As a corollary of the proposition just proved we derive the following result for all automatic linear orderings:

**Theorem 4.5** *The $FC$–rank of every automatic linear order is finite.*

**Proof** Let $\mathcal{L}$ be a linear order and write it as $\sum\{\mathcal{L}_i \mid i \in D\}$ where $\mathcal{D}$ is dense and each $\mathcal{L}_i$ is scattered. We will show that for every $i \in D$ and every $a,b \in L_i$, the $VD$–rank of $[a,b]$ is uniformly bounded. Let $(Q_{\leq}, \iota_{\leq}, \Delta_{\leq}, F_{\leq})$ be a deterministic 2–tape automaton recognising the ordering of $\mathcal{L}$. Let $(Q_A, \iota_A, \Delta_A, F_A)$ be a deterministic 3–tape automaton recognising the definable relation $\{(x,z,y) \mid x \leq z \leq y\}$. Now consider an interval $[a,b]$ of $\mathcal{L}_i$ for some $i \in D$. The proof of the previous theorem ensures that the $VD$–rank of the scattered interval $[a,b]$ is at most $e$, where the constant $e$ does not depend on $[a,b]$ but only on $|Q_A|$ and $|Q_{\leq}|$. Therefore the $VD$–rank of $[a,b]$ interval is at most $e$. Hence $VD(\mathcal{L}_i) \leq e$ for every $i \in D$ and so $FC(\mathcal{L}) \leq e$. □

**Corollary 4.6** [3] *An ordinal $\alpha$ is automatically presentable if and only if $\alpha < \omega^\omega$.*

**Proof** Suppose $\alpha$ is an automatically presentable ordinal. Then by Theorem 4.5 it has finite $FC$–rank and so by Example 3.5 $\alpha < \omega^\omega$ as required.

Given $\alpha < \omega^\omega$ there exists (a least) $n < \omega$ such that $\alpha < \omega^n$. But $\omega^n$ is automatically presentable. Say $(W, <)$ is an automatic presentation, and let $p \in W$ be the string corresponding to $\alpha$. Then the ordinal $\alpha$ is definable as the induced order on the domain $\{x \in W \mid x < p + 1\}$. So $\alpha$ is automatically presentable as required. □

**Proposition 4.7** *It is decidable whether or not an automatic linear order $\mathcal{L}$ is scattered. If it is not scattered then a regular dense subordering is effectively computable from a presentation for $\mathcal{L}$.*

**Proof** Let $\mathcal{L}$ be an automatic order. The proof of Theorem 4.5 says that a bound $e$ on the $FC$–rank of $\mathcal{L}$ is computable given automata for the order and the interval relation. The condensation binary relation $c_{FC}$ is definable in $\mathcal{L}$ since $c_{FC}(x) = c_{FC}(y)$ if and only if $[x, y]$ is finite. Since the ordering on $c_{FC}[\mathcal{L}]$ is also definable from $\mathcal{L}$ (see Definition 3.7) the linear orders $c^i_{FC}[\mathcal{L}]$ are definable for every $i \in \mathbb{N}$, and hence automatic. So consider $c^e_{FC}[\mathcal{L}]$. By Example 3.9 it is isomorphic to **1** if and only if $\mathcal{L}$ is scattered. So using the decidability of the theory of $c^e_{FC}[\mathcal{L}]$, check this with the sentence $\exists x, y\, x < y$. In case $c^e_{FC}[\mathcal{L}]$ is not the singleton it must be a countably infinite dense ordering. One may view $c^e_{FC}$ as an automatic equivalence relation on $\mathcal{L}$ (the $c^e_{FC}(x)$'s partition $\mathcal{L}$), and so the $<_{llex}$-smallest representatives from every equivalence class forms a dense subordering of $\mathcal{L}$ that is a regular subset of $L$. □

# 5 Decidability results for automatic ordinals

Theorem 4.5 can now be applied to prove decidability results for automatic ordinals.

**Proposition 5.1** *Let $\mathcal{L} = (L, <)$ be an automatic structure. It is decidable whether $\mathcal{L}$ is isomorphic to an ordinal.*

**Proof** First check that $<$ linearly orders $L$, by testing whether $\mathcal{L}$ is reflexive, transitive and anti-symmetric – all first order axioms. Although the axiom for being a well-order is not first order expressible, see for instance [14, Theorem 13.13], the following algorithm can be used.

1. **Input** the presentation $(L, <)$ of $\mathcal{L}$.

2. Let $D = L$.

3. **While** $(D, <)$ is not dense and $\forall x \in D\, \omega^*$ does not embed in the interval $c(x)$

   **Do** Replace $(D, <)$ by a presentation for $c[\mathcal{D}]$.

4. **End While**

5. If $\mathcal{D}$ is isomorphic to **1** then **Output $\mathcal{L}$ is an ordinal**,
   else **Output $\mathcal{L}$ is not an ordinal**.

Every step in the algorithm is computable. Indeed the equivalence relation on pairs $(x, y)$ satisfying $c(x) = c(y)$ is definable as $(\neg\exists^\infty z)\,[x < z < y]$. So a presentation for $c[\mathcal{D}]$ is computed by factoring $\mathcal{D}$ by $c$. The while test is expressible as

$$(\forall x \neq y)\,(\exists z)\,[x < z < y]$$

and

$$(\forall x)(\neg\exists^\infty y)\,(c(x) = c(y) \wedge y < x).$$

The final test is expressible by $(\exists x)\,(\forall y)\,[x = y]$.

Since the $FC$–rank of $\mathcal{L}$ is finite, say $k$, the algorithm terminates after at most $k+1$ many while-loop tests. If $\mathcal{L}$ is an ordinal then using the properties of $c$, $c[\mathcal{L}]$ is an ordinal and for every $x \in L$, $c(x)$ is either finite or isomorphic to $\omega$. By induction on $k$, for every $0 \leq i \leq k$, $c^i[\mathcal{L}]$ passes the $(i+1)$'th while-test. The resulting order $\mathcal{D} = c^k[\mathcal{L}]$ is isomorphic to **1** as required.

If $\mathcal{L}$ is not an ordinal then there exists an infinite decreasing sequence of elements. Suppose there exists such a sequence $x_1 > x_2 > x_3 \ldots$ and an $n_0 \in \mathbb{N}$ such that for all $i \geq n_0$, $c(x_i) = c(x_{n_0})$. Then the while-test fails the first time it is executed and the resulting order $\mathcal{D} = \mathcal{L}$ is not isomorphic to the ordinal **1**. If there is no such sequence $(x_i)$ and $n_0$ then there exists a sequence, say $y_1 > y_2 > y_3 > \ldots$ such that $c(y_{i+1}) \ll c(y_i)$ for all $i \in \mathbb{N}$; this is an infinite decreasing sequence of elements in $c[\mathcal{L}]$. Continue inductively in this way with $c[\mathcal{L}]$ in place of $\mathcal{L}$. Suppose the while-test fails the $m$'th time where $1 \leq m \leq k$. If it fails because there is some $x \in c^{m-1}[\mathcal{L}]$ for which $\omega^\star$ embeds in $c(x)$ then $\mathcal{D} = c^{m-1}[\mathcal{L}]$ is infinite and so not isomorphic to **1**. If this does not occur, then the while-test fails the $k+1$'st time because $\mathcal{D} = c^k[\mathcal{L}]$ is dense. But then there is a sequence $z_1 > z_2 > z_3 > \ldots$ with $c^k(z_{i+1}) \ll c^k(z_i)$ for every $i \in \mathbb{N}$, and so $\mathcal{D}$ not isomorphic to **1**. □

We now show that the isomorphism problem for automatic ordinals is decidable. Contrast this with the fact that the isomorphism problem for computable ordinals is $\Pi^1_1$-complete. Recall that by Cantor's Normal Form Theorem if $\alpha$ is an ordinal then it can be uniquely decomposed as $\omega^{\alpha_1}n_1 + \omega^{\alpha_2}n_2 + \ldots + \omega^{\alpha_k}n_k$, where $\alpha_1, \alpha_2, \ldots, \alpha_k$ are ordinals satisfying $\alpha_1 > \alpha_2 > \ldots > \alpha_k$ and $k, n_1, n_2, \ldots, n_k$ are natural numbers. The proof of deciding the isomorphism problem for automatic ordinals is based on the fact that the Cantor's normal form can be extracted from automatic presentations of ordinals.

**Theorem 5.2** *If $\alpha$ is an automatic ordinal then its normal form is computable from an automatic presentation of $\alpha$.*

**Proof** Let $(R, \leq_{ord})$ be an automatic presentation over $\Sigma$ of $\alpha$. Recall that the unknown ordinal is of the form $\alpha = \omega^m n_m + \omega^{m-1}n_{m-1} + \ldots + \omega^2 n_2 + \omega n_1 + n_0$ where $m, n_m, n_{m-1}, \ldots, n_1, n_0$ are natural numbers. Now one can compute the values of the numbers $m, n_0, n_1, \ldots$ by the following algorithm.

1. **Input** the presentation $(R, \leq_{ord})$.

2. Let $D = R$, $m = 0$, $n_m = 0$.

3. **While** $D \neq \emptyset$ **Do**

4. **If** $D$ has a maximum $u$

   **Then** Let $n_m = n_m + 1$, let $D = D - \{u\}$.

   **Else** Let $L \subseteq D$ be the set of limit ordinals in $D$; that is $L$ is the set of all $x \in D$ with no immediate predecessor in $D$. Replace $D$ by $L$, let $m = m + 1$, let $n_m = 0$.

5. **End While**

6. **Output** the formula

$$\omega^m n_m + \omega^{m-1} n_{m-1} + \ldots + \omega^2 n_2 + \omega n_1 + n_0$$

   using the current values of $m, n_0, \ldots, n_m$.

Since the first order theory of an automatic structure is decidable, each step in the algorithm is computable. Removing the maximal element from $D$ reduces the ordinal represented of $D$ by 1 while the corresponding $n_m$ is increased by 1. Replacing $D$ by the set of its limit ordinals is like dividing the ordinal represented by $D$ by $\omega$; the set of limit ordinals (including 0) strictly below $\omega^m a_m + \ldots + \omega^1 a_1$ has order type $\omega^{m-1} a_m + \ldots + \omega^1 a_2 + a_1$. So the next coefficient can start to be computed. Based on this it is easy to verify that the algorithm computes the coefficients $n_0, n_1, \ldots$ in this order. The algorithm eventually terminates since $m$ is bounded by the finite bound on the $VD$–rank of the ordinal. □

The following is an immediate corollary.

**Theorem 5.3** *The isomorphism problem for automatic ordinals is decidable.*

Compare this with the fact that the isomorphism problem for automatic structures and even permutation structures [1, 7] is not decidable.

**Problem 5.4** *Is the isomorphism problem for automatic linear orders decidable ?*

# 6 Automatic tree preliminaries

The remaining sections deal with trees viewed as partial orders. Theorems 7.6 and 7.9 give a necessary condition for certain trees to be automatic. The condition is similar to that for linear orders and says that the Cantor-Bendixson rank (Definition 7.1) of the tree be finite.

A *tree* $\mathcal{T} = (T, \preceq)$ is a partial order that has a least element $r$, called the root, and in which $\{y \in T \mid y \preceq x\}$ is a finite linear order for each $x \in T$. So we think of trees as growing upwards. Write $x \| y$ if $x \not\preceq y$ and $y \not\preceq x$. A partial order $(T, \preceq)$ is a *forest* if there is a partition of the domain $T = \cup T_i$ such that every $(T_i, \preceq)$ is a tree. The subtree rooted at $x$, written $\mathcal{T}(x)$, has domain $T(x) = \{y \in T \mid x \preceq y\}$ with order $\prec$ restricted to this domain. The set $S(x)$ of immediate successors of $x$ is defined as

$$S(x) = \{y \in \mathcal{T} \mid x \prec y \wedge (\forall z)[x \preceq z \preceq y \rightarrow (z = x \vee z = y)]\}.$$

A tree $\mathcal{T}$ is *finitely branching* if $S(x)$ is finite for each $x \in T$. A *path* of a tree $(T, \preceq)$ is a subset $P \subseteq T$ which is linearly ordered (by $\preceq$), closed downward (that is, whenever $y \in P$ and $x \preceq y$ then $x \in P$), and maximal (under set-theoretic inclusion) with these properties. A path with finitely many elements is called a *finite path*; otherwise it is called an *infinite path*.

Recall that $<_{llex}$ is the length lexicographic order on $\Sigma^\star$ defined as $x <_{llex} y$ if either $|x| < |y|$ or $|x| = |y|$ but $x$ lexicographic before $y$. For example, $\epsilon <_{llex} 0 <_{llex} 1 <_{llex} 00 <_{llex} 01 <_{llex} \dots$ in the case that $\Sigma = \{0, 1\}$. Thus if $\mathcal{T}$ is an automatic tree with $T \subset \Sigma^\star$ then the length-lexicographic order on $\Sigma^\star$ is inherited by each set $S(x)$. This permits one to talk about the first, second, third, ... successor of $x$.

# 7 Ranks of automatic trees

Our approach to proving facts about trees is to associate a linear order with a tree, in such a way that the tree is automatic if and only if the linear order is automatic. Then by Theorem 4.5 the linear order has finite rank which it turns out implies that the rank of the tree is finite. More precisely, in this section it is shown that every automatic tree has finite Cantor-Bendixson Rank.

Given a tree $\mathcal{T}$, define a subset of $T$ as consisting of those nodes $x \in T$ with the property that there exist at least two distinct infinite paths in the subtree of $\mathcal{T}$ rooted at $x$. It follows from the definitions that this sub-partial order, $d(\mathcal{T})$, is in fact a subtree of $\mathcal{T}$.

For each ordinal $\alpha$ define the iterated operation $d^\alpha(\mathcal{T})$ inductively as follows.

(1) $d^0(\mathcal{T}) = \mathcal{T}$.

(2) $d^{\alpha+1}(\mathcal{T})$ is $d(d^\alpha(\mathcal{T}))$.

(3) If $\alpha$ is a limit ordinal, then $d^\alpha(\mathcal{T})$ is $\cap_{\beta<\alpha} d^\beta(\mathcal{T})$.

**Definition 7.1** [8] The *Cantor-Bendixson Rank* of a tree $\mathcal{T}$, written $CB(\mathcal{T})$, is the least ordinal $\alpha$ such that $d^\alpha(\mathcal{T}) = d^{\alpha+1}(\mathcal{T})$.

**Remark 7.2** The Cantor-Bendixson Rank of an arbitrary topological space $X$ is defined as above, using $D$ given as $D\,X = \{P \in X \mid p \text{ is not isolated}\}$ instead of $d$. Recall that $P$ is isolated if $\{P\}$ is an open set. So given a tree $\mathcal{T} = (T, \preceq)$, consider the following topological space. The set of elements are the infinite paths in $\mathcal{T}$, written $[\mathcal{T}]$. For $P \in [\mathcal{T}]$ and $x \in T$ write $x \prec P$ if $x \in P$ and say that $x$ is on $P$. The basic open sets are of the form $\{P \in [\mathcal{T}] \mid x \prec P\}$ for every $x \in T$. Then the Cantor-Bendixson Rank of this topological space, $CB[\mathcal{T}]$, is just the least ordinal $\alpha$ such that $D^{\alpha+1}[\mathcal{T}] = D^\alpha[\mathcal{T}]$. Given an infinite path $P$ of $\mathcal{T}$, the following statements are equivalent:

- There is a node $x \prec P$ such that $P$ is the only infinite path of $\mathcal{T}$ going through $x$;
- $P \notin D(\mathcal{T})$;
- There is a $x \prec P$ with $x \notin d(\mathcal{T})$.

It follows that $D[\mathcal{T}]$ consists of exactly the infinite paths of $d(\mathcal{T})$. It can be proven by transfinite induction that also

$$D^\alpha[\mathcal{T}] = [d^\alpha(\mathcal{T})].$$

Assume now that $\alpha = CB[\mathcal{T}]$. Then $d^\alpha(\mathcal{T})$ and $d^\beta(\mathcal{T})$ contain the same infinite paths for all $\beta > \alpha$, but $d^\alpha(\mathcal{T})$ might contain some nodes which are not on any infinite paths and therefore not contained in $d^{\alpha+1}(\mathcal{T})$. Thus the two CB-ranks might differ, but they differ at most by 1:

$$CB[\mathcal{T}] \leq CB(\mathcal{T}) \leq CB[\mathcal{T}] + 1.$$

A witness $\mathcal{T}$ with $CB[\mathcal{T}] \neq CB(\mathcal{T})$ is the tree where the domain consists of the root 0 and, for every $n > 0$, the strings $01^{a_1}01^{a_2}0\ldots1^{a_n}0$ with $a_1 \geq a_2 \geq \ldots \geq a_n$; the ordering is the prefix-relation $\preceq$ restricted to this domain. One has for every node $01^{a_1}01^{a_2}0\ldots1^{a_n}0 \in T$ that $01^{a_1}01^{a_2}0\ldots1^{a_n}0 \in d^m(\mathcal{T}) \Leftrightarrow a_n \geq m$. So $d^\omega = \{0\}$. It follows that $CB[\mathcal{T}] = \omega$ by $D^\omega(\mathcal{T}) = \emptyset$ while $CB(\mathcal{T}) = \omega + 1$ by $d^{\omega+1}(\mathcal{T}) = \emptyset \neq d^\omega(\mathcal{T})$. This witness is also robust to small changes in the definition of $d$. If one, for example, takes $d(\mathcal{T})$ to contain exactly those nodes which are on infinitely many infinite paths of $\mathcal{T}$, then the resulting trees $d^\alpha(\mathcal{T})$ and derived CB-ranks are the same.

Here are some basic properties of $CB$–rank that will be used without reference.

**Property 7.3** *If $\mathcal{T}$ is a tree with $CB(\mathcal{T}) = \alpha$ then*

(1) *$\alpha$ is a countable ordinal.*

(2) *If $d^\alpha(\mathcal{T}) \neq \emptyset$ then $d^\alpha(\mathcal{T})$ and $\mathcal{T}$ contain uncountably many infinite paths.*

(3) *If $d^\alpha(\mathcal{T}) = \emptyset$ then $\mathcal{T}$ contains only countably many infinite paths. Furthermore, $\alpha$ is either 0 or a successor ordinal.*

**Proof** For each $\beta$ let $x_\beta \in d^\beta(\mathcal{T}) \setminus d^{\beta+1}(\mathcal{T})$. Since $T$ is countable, and $\alpha \neq \beta$ implies that $x_\alpha \neq x_\beta$, the set of ordinals $\beta$ such that $d^\beta(\mathcal{T}) \setminus d^{\beta+1}(\mathcal{T}) \neq \emptyset$ is also countable. Hence its least upper bound, a countable ordinal, say $\alpha$, is $CB(\mathcal{T})$. This proves (1).

If $d^\alpha(\mathcal{T})$ is not the empty tree, then for every $x \in d^\alpha(\mathcal{T})$ there exist $y, z \in d^\alpha(\mathcal{T})$ with $x \prec y, z$ and $y \| z$. In particular the full binary tree $(\{0,1\}^\star, \prec_p)$ embeds in $d^\alpha(\mathcal{T})$. Since $d^\alpha(\mathcal{T})$ is a subset of $\mathcal{T}$, the full binary tree also embeds in $\mathcal{T}$. This proves (2).

If $d^\alpha(\mathcal{T})$ is the empty tree, then one shows that $\mathcal{T}$ has only countably many infinite paths as follows: For every infinite path $P$ of $\mathcal{T}$ there is a minimum ordinal $\beta_p \leq \alpha$ such that $P \not\subseteq d^{\beta_P}(\mathcal{T})$. Furthermore, there is a node $x_P$ in $P$ such that $x_P \notin d^{\beta_P}(\mathcal{T})$. Since $x_P \in d^\gamma(\mathcal{T})$ for all $\gamma < \beta_P$, it holds that $\beta_P$ is a successor ordinal $\delta + 1$. Furthermore, $P$ is the only infinite path of $d^\delta(\mathcal{T})$ which contains $x_P$. Thus the mapping $P \to (x_P, \beta_P)$ of the infinite paths of $\mathcal{T}$ to pairs of nodes and successor ordinals up to $\alpha$ is one-one. Since the range of this mapping is countable, so is its domain. Suppose $\alpha > 0$ is a limit ordinal. Furthermore, $d^\alpha(\mathcal{T})$ is non-empty since the root of $\mathcal{T}$ is in $d^\gamma(\mathcal{T})$ for every $\gamma < \alpha$. So $\alpha$ is also a successor ordinal. This proves (3). □

The following lemma will be used in the next theorem. Recall that $\mathcal{T}(x)$ is the subtree of $\mathcal{T}$ rooted at $x$.

**Lemma 7.4** *Suppose that $CB(\mathcal{T}) \geq \beta+1$ and $\mathcal{T}$ has countably many infinite paths.*

(1) *There is an $x' \in T$ such that $CB(\mathcal{T}(x')) = \beta+1$.*

(2) *If $\beta > 1$ then at least one of the following holds:*

- *there exists $x' \in T$, $x' \neq x$, $CB(\mathcal{T}(x')) = \beta+1$,*
- *there exist $y, y' \in T$ with $y \| y'$, both on an infinite path of $\mathcal{T}$ and $CB(\mathcal{T}(y)) = \beta$ and $CB(\mathcal{T}(y')) = \beta$.*

(3) *If $P$ is an infinite path of $T$ and the $CB$–rank of $\mathcal{T}(x)$ is $\beta+1$ for almost all $x$ on $P$, then for every node $x$ on $P$ and every $\gamma < \beta$ there is a node $y_{\gamma,x} \succeq x$ such that $y_{\gamma,x} \in T$ and $\mathcal{T}(y)$ has $CB$–rank $\gamma+1$. None of the nodes $y_{\gamma,x}$ is on $P$.*

**Proof** First note that if $\mathcal{T}$ has countably many infinite paths and $CB(\mathcal{T}) = \alpha$ then $d^{\alpha}(\mathcal{T}) = \emptyset$ and so for every $y \in T$ there is some $\gamma \leq \alpha$ such that $y \notin d^{\gamma}(\mathcal{T})$. The least such $\gamma$ is equal to the $CB$–rank of $\mathcal{T}(y)$.

Assume that the $CB$–rank of $\mathcal{T}$ is at least $\beta+1$. Then $d^{\beta}(\mathcal{T})$ is non-empty, since it contains $x$. Furthermore $d^{\beta+1}(\mathcal{T}) \neq d^{\beta}(\mathcal{T})$. So pick a $y$ in $d^{\beta}(\mathcal{T})$ and not in $d^{\beta+1}(\mathcal{T})$. Then $\beta+1$ is the least ordinal such that $y \notin d^{\beta+1}(\mathcal{T}(x))$; hence $CB(\mathcal{T}(y)) = \beta+1$. This proves (1).

Let $y_1, \ldots, y_k$ be the immediate successors of $x$. There are two different cases.

*Case* 1. Suppose some successor, say $y_1$, is in $d^{\beta}(\mathcal{T})$. Then the $CB$–rank of $\mathcal{T}(y')$ is at least $\beta+1$ and so by item (1) there is an $x' \succeq y_1$ such that $x'$ in $T(y')$ and the $CB$–rank of $\mathcal{T}(x')$ is equal to $\beta+1$.

*Case* 2. It holds for $y_1, \ldots, y_k$ that none of them is in $d^{\beta}(\mathcal{T})$. For each $l$ let $\beta_l$ be the $CB$–rank of $\mathcal{T}(y_l)$ and let $\gamma$ be the maximum of all $\beta_l$. The ordinal $\gamma$ is the successor of some other ordinal, say $\gamma = \delta+1$. Note that $\gamma = \beta$ since otherwise $x \notin d^{\beta}(\mathcal{T})$. There are at least two infinite paths $P, P'$ of $d^{\delta}(\mathcal{T})$ since $\beta > 1$ by hypothesis. There are $i, j$ such that $P$ goes through $y_i$ and $P'$ through $y_j$. The $i, j$ must be different since otherwise $y_i \in d^{\beta}(T)$ in contrary to the assumption of Case 2. It holds that both $y_i, y_j$ are in $d^{\delta}(\mathcal{T})$ and not in $d^{\delta+1}(\mathcal{T})$ and the $CB$–ranks of $\mathcal{T}(y_i)$ and $\mathcal{T}(y_j)$ are $\beta$. This completes the proof of (2).

Item (3) follows from the observation that $CB(\mathcal{T}(x)) = \beta+1$ implies that for every $\gamma \leq \beta$ there exists $y \in T(x)$ such that $CB(\mathcal{T}(y)) = \gamma+1$. For otherwise if $\gamma < \beta$ were a counterexample then by item (1) it holds that $CB(\mathcal{T}(y)) \leq \gamma+1$ for every $y \in T(x)$, contradicting the case $y = x$. That none of the nodes are on $p$ follows from the fact that if $x$ is on $p$ then $CB(\mathcal{T}(x)) = \beta+1$. □

For the first result one associates the Kleene-Brouwer ordering with a tree.

**Definition 7.5** [13] Let $(T, \preceq)$ be a tree and $\leq_{llex}$ be the length lexicographic order induced by the presentation of $T$ as a subset of $\Sigma^*$. Let $x, y$ be nodes on $T$. Then $x \leq_{kb} y$ iff either $y \preceq x$ or there are $u, v, w$ such that $v, w \in S(u)$, $v \preceq x$, $w \preceq y$ and $v <_{llex} w$. Write $\mathcal{L}_T$ for the structure $(T, \leq_{kb})$.

In words, $x \leq_{kb} y$ if and only if either $x$ is above $y$ in the tree or $x$ is to the left of $y$ (with respect to $<_{llex}$ restricted to immediate successors). Note that $\leq_{kb}$ linear orders $T$ and $(T, \leq_{kb})$ is first order definable from $(T, \preceq, \leq_{llex})$.

**Theorem 7.6** *The $CB$–rank of an automatic finitely branching tree with countably many infinite paths is finite.*

**Proof** Suppose $\mathcal{T}$ is finitely branching with countably many infinite paths and $CB(\mathcal{T}) = \alpha$. Then $\alpha$ is either 0 or $\beta + 1$ for some ordinal $\beta$. We now prove by induction on $\alpha$ that $\mathcal{L}_T$ is scattered and

$$VD_*(\mathcal{L}_T) \leq \alpha \text{ and if } \alpha > 0 \text{ then } \beta \leq VD_*(\mathcal{L}_T).$$

Consequently if $\mathcal{T}$ is automatic then so is $\mathcal{L}_T$, which by Theorem 4.5 has finite $VD$–rank and hence finite $VD_*$–rank. Then $CB(\mathcal{T})$ must also be finite as required.

We deal with the bases cases first. If $\alpha = 0$ then $\mathcal{T}$ is the empty tree so $\mathcal{L}_T$ is $\mathbf{0}$ which has $VD(\mathcal{L}_T) = VD_*(\mathcal{L}_T) = 0$. If $\alpha = 1$ then $\mathcal{T}$ is non-empty and contains at most one infinite path. By the definition of $<_{kb}$, $\mathcal{L}_T$ has order type $\mathbf{n}$, $\omega^\star$, $\mathbf{n} + \omega^\star$ or $\omega + \omega^\star$, where $n \in \mathbb{N}$. In these cases the $VD_*$–rank of $\mathcal{L}_T$ is either 0 or 1 and so the result holds.

Now consider $\alpha > 1$. For $x \in T$ recall that $\mathcal{T}(x)$ is the subtree of $\mathcal{T}$ rooted at $x$, that is $\{y \in T \mid x \preceq y\}$. Define $X = \{x \in T \mid CB(\mathcal{T}(x)) = \alpha\}$. Then $X \neq \emptyset$ since the root of $\mathcal{T}$ is in $X$. Further $X = d^\beta(\mathcal{T})$. Indeed if $x \in X$ then in particular $x \in d^\beta(\mathcal{T}(x))$ and so $x \in d^\beta(\mathcal{T})$. Conversely if $x \notin X$ then $d^\gamma(\mathcal{T}(x)) = \emptyset$ for some $\gamma \leq \beta$. Hence $x \notin d^\gamma(\mathcal{T})$ and in particular $x \notin d^\beta(\mathcal{T})$. So $X$ is a tree and $CB(X) \leq 1$. But since $X$ is non-empty $CB(X) = 1$ and so by definition of $d$ and using the fact that $d(X) = \emptyset$, $X$ contains at most one infinite path, and so $\mathcal{L}_X$ has one of the forms listed in the case $\alpha = 1$.

For $x \in X$ define a possibly empty set $M_x$ as the union of the $\mathcal{T}(y)$'s where $y$ is $\prec$-minimal with the property

$$x <_{kb} y \wedge (\forall x' \in X)\, [x' \leq_{kb} y \rightarrow x' \leq_{kb} x].$$

For a given $x \in X$ these $y$'s are pairwise incomparable by minimality. In other words $M_x$ consists of all those $z \in T$ such that the half open interval $\{z' : x <_{kb} z' \leq_{kb} z\}$ in $\mathcal{L}_T$ contains no node of $X$. And conversely if $z \in T \setminus X$ and $\{z' \in T \mid z' <_{kb} z\} \cap X$ has a maximum element, say $x$, then $z \in M_x$. Now note that each such $\mathcal{T}(y)$ has $CB$–rank at most $\beta$ since $y \notin X$.

*Case* 1. Suppose $X$ is finite. Then $\mathcal{X}$ has a maximal element, say $x$. Then by Lemma 7.4 part 2($b$), $CB(\mathcal{T}(x)) = \beta + 1$ implies that there are $\mathcal{T}(y)$ and $\mathcal{T}(y')$ in $\mathcal{M}_x$ with $CB(\mathcal{T}(y)) = CB(\mathcal{T}(y')) = \beta$. So $\mathcal{M}_x$ is a forest with finitely many trees $\mathcal{T}(y)$, of which at least two have $CB$–rank exactly $\beta$. So by the inductive hypothesis $\mathcal{L}_{M_x}$ is a finite sum of scattered linear orders $\mathcal{L}_{T(y)}$ of $VD_*$–rank at most $\beta$, of which at least two have $VD_*$–rank exactly $\beta$. Hence $\mathcal{L}_{M_x}$, being a finite sum of these $\mathcal{T}(y)$'s since $\mathcal{T}$ is finitely branching, has $VD_*$–rank exactly $\beta$. Now

$$\mathcal{L}_T = \sum \{z + \mathcal{L}_{M_z} \mid z \in X\},$$

and so $\mathcal{L}_T$ is scattered and has $VD_*$–rank $\beta$ as required.

*Case* 2. Suppose $X$ is infinite. Let $P = (x_i)_{i \in \mathbb{N}}$ be the infinite path in $\mathcal{X}$. For $x_i$ define $L_i$, respectively $R_i$, as the possibly empty forest consisting of trees $\mathcal{T}(y)$ where $y \in S(x_i)$ and $y <_{kb} x_{i+1}$, respectively $x_{i+1} <_{kb} y$. Since $\mathcal{T}$ is finitely branching each $L_i$ and $R_i$ contains finitely many trees $\mathcal{T}(y)$, each of which has $CB$–rank exactly $\beta + 1$ if

$y \in X$ and at most $\beta$ otherwise. For the former, $\mathcal{L}_{T(y)}$ is scattered and has $VD_*$–rank $\beta$ by case 1. For the latter, $\mathcal{L}_{T(y)}$ is scattered and has $VD_*$–rank at most $\beta$ by the inductive hypothesis. Now by Lemma 7.4 part 3, since $CB(\mathcal{T}(x_i)) = \beta + 1$ for every $i \in \mathbb{N}$, it must be the case that the supremum of $CB(\mathcal{T}(y))$ where the $\mathcal{T}(y)$'s are from $L_i$ and $R_i$, $i \in \mathbb{N}$, is at least $\beta$. Applying the inductive hypothesis to the latter $\mathcal{T}(y)$'s and possible use of the former case, the supremum of the $VD_*$–ranks of the $\mathcal{L}_{T(y)}$'s is at least and hence exactly $\beta$. So the supremum of the $VD_*$–ranks of the $\mathcal{L}_{L_i}$'s and $\mathcal{L}_{R_i}$'s, each being a finite sum of $\mathcal{L}_{T(y)}$'s, is $\beta$. Furthermore,

$$\mathcal{L}_T = \mathcal{L}_{L_0} + \mathcal{L}_{L_1} + \mathcal{L}_{L_2} + \ldots + x_2 + \mathcal{L}_{R_1} + x_1 + \mathcal{L}_{R_0} + x_0$$

Note that $\mathcal{L}_T$ being a scattered sum of scattered linear orders is scattered itself. Also by definition of $VD_*$ and using that $\beta > 0$ each term $\mathcal{L}_{L_i}$ and $(\mathcal{L}_{R_i} + x_i)$ is a finite sum of orders of $VD$–rank equal to $\beta$. Hence the $VD$–rank of $\mathcal{L}_T$ is at least $\beta + 1$. Moreover since infinitely many of the linear orders amongst the $\mathcal{L}_{L_i}$'s and $\mathcal{L}_{R_i}$'s are finite sums of linear orders of $VD$–rank exactly $\beta$, it holds that $\mathcal{L}_T$ has $VD$–rank exactly $\beta+1$. So $\mathcal{L}_T$ also has $VD_*$–rank $\beta + 1$ as required. □

Define the extendible part $E(\mathcal{T})$ of $\mathcal{T}$ as those $x \in T$ that are on some infinite path of $T$. Say that $\mathcal{T}$ is *pruned* if $\mathcal{T} = E(\mathcal{T})$. Then $d(\mathcal{T})$ is the subtree of $\mathcal{T}$ restricted to domain $\{x \in E(\mathcal{T}) \mid \exists z, z' \in E(\mathcal{T}), z, z' \succ x \text{ and } z \| z'\}$. Note that if $\mathcal{T}$ is finitely branching then $d(\mathcal{T})$ is definable in $\mathcal{T}$ since $E(\mathcal{T})$ is then equivalent to $\{x \in T \mid \exists^\infty y, x \prec y\}$. A tree $\mathcal{P}$ is *perfect* if $\mathcal{P} = d(\mathcal{P})$. In other words it satisfies the condition

$$\forall x \in P, \exists z, z' \in E(\mathcal{P})\, x \prec z, z' \wedge z \| z'.$$

In particular a perfect tree is pruned and is either empty or contains uncountably many infinite paths. Given $\mathcal{T}$ define $P_T$ as the set of all $x \in T$ such that $\mathcal{T}(x)$ contains a perfect subtree. This immediately implies that $P_T$ is downward closed and so $\mathcal{P}_T$ is a (possibly empty) tree. In fact $\mathcal{P}_T$ is perfect since if $\mathcal{T}(x)$ contains a perfect subtree then there are $z, z' \in E(\mathcal{T}(x))$ with $z \| z'$.

**Theorem 7.7** *The $CB$–rank of every finitely branching automatic tree is finite.*

**Proof** Suppose $\mathcal{T}$ is a finitely branching automatic tree. Then $P_T$, being definable in $\mathcal{T}$, is a regular subset of $T$. Let $C = T \setminus P_T$ and note that, in general, $\mathcal{C}$ is a forest. For every $x \in C$ that is $\prec$-minimal, let $\alpha_x = CB(\mathcal{T}(x))$. Note that $\mathcal{T}(x)$ is a subtree of $\mathcal{C}$ and so has countably many infinite paths. Let $\alpha$ be the supremum of the $\alpha_x$'s. Then $d^\alpha(\mathcal{T}(x)) = \emptyset$ for every $x \in C$, and so $C \cap d^\alpha(\mathcal{T}) = \emptyset$. That is $d^\alpha(\mathcal{T}) \subset P_T$. Conversely if $x \in P_T$ then $x \in d^\beta(\mathcal{T}(x))$ for every ordinal $\beta$. So $P_T \subset d^\alpha(\mathcal{T})$. Hence $\mathcal{P}_T = d^\alpha(\mathcal{T})$. Now since $d(\mathcal{P}_T) = \mathcal{P}_T$, one has that $CB(\mathcal{T}) \leq \alpha$. But $CB(\mathcal{T})$ can be no smaller than $\alpha$ and so $CB(\mathcal{T}) = \alpha$. It is sufficient to prove that $\alpha$ is finite. First note that by the proof of the Theorem 7.6, it holds that $\mathcal{L}_C$ is scattered (since it has countably many infinite paths) and $\alpha_x \leq VD_*(\mathcal{L}_{T(x)}) + 1$; the latter being finite by Theorem 4.5. □

Next we remove the condition that the tree be finitely branching.

**Definition 7.8** Given a tree $(T, \preceq)$, define a partial order $x \preceq' y$ on $T$ by

$$x \preceq y \vee \exists v, w \in T\, (x, w \in S(v) \wedge x \leq_{llex} w \wedge w \preceq y);$$

where $\leq_{llex}$ the length lexicographic order and $S(v)$ the set of immediate successors of $v$ with respect to $\preceq$.

Then $(T, \preceq')$ is indeed a tree which we denote by $\mathcal{T}'$.

**Theorem 7.9** *The CB–rank of an automatic tree is finite.*

**Proof** Let $\mathcal{T} = (T, \preceq)$ be given and $S(x)$ be the set of immediate successors of an $x \in T$. Furthermore, let $s(x)$ be the length-lexicographically least element of $S(x)$ for the case $S(x) \neq \emptyset$ and let $s(x) = u$ for a default value $u \notin T$ if $S(x) = \emptyset$.

Now let $\mathcal{T}' = (T, \preceq')$ as in Definition 7.8. Note that $\preceq'$ extends $\preceq$. For $x \in T$ let $S'(x)$ be the set of successors with respect to $\preceq'$. $S'(x)$ contains $s(x)$ whenever $s(x) \neq u$ and the length-lexicographically next sibling $y$ of $x$ with respect to $\preceq$ whenever this $y$ exists. Recall that $y$ is a sibling of $x$ with respect to $\preceq$ if there is a node $z$ with $x, y \in S(z)$. Hence $\mathcal{T}' = (T, \preceq')$ is a finitely branching tree.

Let $U$ and $U'$ be the sets of infinite paths of $(T, \preceq)$ and $(T, \preceq')$, respectively. Since every infinite path of $\mathcal{T}$ generates an infinite path of $\mathcal{T}'$, there is a one-one continuous mapping $q$ from $U$ to $U'$. This mapping satisfies for all $P \in U$ and all $x \in T$: $x \in P$ iff $s(x) \in q(P)$. Furthermore, $U'$ contains besides the paths of the form $q(P)$ for some $P \in U$ also the paths generated by those sets $S(x)$ where $S(x)$ is infinite. Since the quantity of these additional paths is countable one has the following equivalence for all $x$: $\{P \in U : x \in P\}$ is uncountable iff $\{P' \in U' : s(x) \in P'\}$ is uncountable.

Now one shows by induction over $n$ that the following implication holds for all $x \in T$ with $s(x) \neq u$ and $n \in \mathbb{N}$: $x \in d^n(\mathcal{T}) \Rightarrow s(x) \in d^n(\mathcal{T}')$. The property clearly holds for $n = 0$. Now assume the inductive hypothesis for $n$ and consider any $x \in d^{n+1}(\mathcal{T})$. There are two distinct infinite paths $P, Q \in U$ such that $x \in P \cap Q$ and $P \cup Q \subseteq d^n(\mathcal{T})$. It follows that $s(x) \in q(P) \cap q(Q)$. By induction hypothesis and by $q$ being one-one, $s(x)$ is member of the two distinct infinite paths $q(P), q(Q)$ of $d^n(\mathcal{T}')$ and thus $s(x) \in d^{n+1}(\mathcal{T}')$. This completes the proof of this property.

By Theorem 7.7, there is a natural number $n$ such that $d^n(\mathcal{T}')$ contains exactly those nodes of the form $s(x)$ which are in uncountably many members of $U'$. Then all $x \in d^n(\mathcal{T})$ satisfy that $x$ is in uncountably members of $U$. On the other hand, every $x$ being in uncountably members of $U$ is in $d^n(\mathcal{T})$. So $d^n(\mathcal{T})$ contains exactly the nodes $x$ which are in uncountably many members of $U$ and $d^{n+1}(\mathcal{T}) = d^n(\mathcal{T})$. The CB-rank of $\mathcal{T}$ is at most $n$. $\square$

# 8 Automatic versions of König's Lemma

König's Lemma says that every infinite finitely branching tree has at least one infinite path. This section consists of automatic versions of this result. If one considers Turing machines instead of finite automata there are trees which have infinite paths, but no hyperarithmetic one, and in particular no computable infinite path. Furthermore even finitely branching trees might have infinite paths but none of them be computable. In contrast to this the following results state that every automatic tree, not necessarily finitely branching, either has a regular infinite path or does not have an infinite path at all.

**Proposition 8.1** *It is decidable whether an automatic tree has an infinite path.*

**Proof** Let $(T, \preceq)$ be an automatic tree and recall that $(T, <_{kb})$ is an automatic linear order. By Proposition 5.1 it is decidable whether this order is isomorphic to an ordinal. And this is the case if and only if $(T, \preceq)$ has no infinite path. To prove this last statement recall that a linear order is isomorphic to an ordinal if and only if it has no infinite decreasing chain. So suppose $(T, \preceq)$ has an infinite path $x_1 \prec x_2 \prec x_3 \ldots$. Then $x_1 >_{kb} x_2 >_{kb} x_3 \ldots$ is an infinite decreasing chain in $(T, \leq_{kb})$, and so $(T, \leq_k b)$ is not isomorphic to an ordinal. Conversely, suppose $(T, <_{kb})$ is not isomorphic to an ordinal and let $x_1 >_{kb} x_2 >_{kb} x_3 \ldots$ be an infinite decreasing chain. We define an infinite path $(p_i)$ of $(T, \prec)$ as follows.

1. Let $i = 1$ and $j = 1$.
2. **Repeat**
   (a) Define $p_i = x_j$.
   (b) Replace $j$ with the smallest $k > j$ for which there is a $u \in S(p_i)$ with $u \preceq x_l$ for every $l \geq k$.
   (c) Replace $i$ with $i + 1$.
3. **End Repeat**

If such a $k$ exists in step $2(b)$ of every stage of the repeat loop, then the resulting sequence $(p_i)$ is an infinite path in $(T, \preceq)$. So suppose that the algorithm has computed $p_1, p_2, \ldots, p_n$ with $p_1 \prec p_2 \prec \ldots \prec p_n$. So $i = n$ and $j \in \mathbb{N}$. For every $m > j$ define $u(x_m)$ as the immediate successor of $p_i$ that is $\preceq x_m$. Then this sequence satisfies $u(x_m) \geq_{llex} u(x_{m+1}) \geq_{llex} u(x_{m+2}) \geq_{llex} \ldots$ since $x_m <_{kb} x_{m+1} <_{kb} x_{m+2} <_{kb} \ldots$. But since $\leq_{llex}$ is isomorphic to an ordinal (of type $\omega$) it can not have an infinite decreasing sequence. Thus the sequence is eventually constant; that is, there is a (smallest) $k > j$ such that for every $l \geq k$ one has $u(x_k) = u(x_l) \preceq x_l$ as required. □

## 8.1 Finitely branching automatic trees

Recall that an infinite tree is *pruned* if every element is on some infinite path. Note that if $\mathcal{T}$ is finitely branching then the set of elements $E(\mathcal{T})$ above which there are infinitely many elements is definable as $\{x \in T \mid (\exists^\infty y)\, x \preceq y\}$. Hence we can restrict attention to the subtree on domain $E(\mathcal{T})$. Indeed, if $x \in E(\mathcal{T})$ then by König's Lemma it is on an infinite path. Conversely if $x \notin E(\mathcal{T})$ then there are only finitely many elements above it (in $\mathcal{T}$) and so it is not on an infinite path. Hence the subtree $(E(\mathcal{T}), \preceq)$ is pruned and has an infinite path iff and only if $\mathcal{T}$ has an infinite path. Further if $\mathcal{T}$ is automatic then so is $E(\mathcal{T})$. This allows one to assume without loss of generality that a *finitely branching* automatic tree is already pruned.

**Theorem 8.2 (Automatic König's Lemma, Version** 1) *If $\mathcal{T} = (T, \preceq)$ is an infinite finitely branching automatic tree then it has a regular infinite path. That is, there exists a regular set $P \subset T$ so that $P$ is an infinite path of $\mathcal{T}$.*

**Proof** By the previous remark suppose that $\mathcal{T}$ is pruned. Recall that the length-lexicographic order $<_{llex}$ on $\Sigma^*$ is automatic and therefore one can extend the presentation of $\mathcal{T}$ to include $<_{llex}$. Now define the leftmost infinite path $P$ with respect to the length-lexicographic order of the successors of any node. $P$ contains those nodes $x$ for which every $y \prec x$ satisfies that $\forall z, z' \in S(y)\, [z \preceq x \Rightarrow z <_{llex} z']$. This means, that the unique node $z \in S(y)$ which is below $x$ is just the length-lexicographically least element of $S(y)$. Since the length-lexicographic ordering of $\Sigma^*$ is a well-ordering (of type $\omega$), this minimum always exists.

We briefly check that $P$ is an infinite path. Firstly $P$ is closed downward. Indeed, given $x \in P$, let $a \preceq x$. Then for every $y \prec a$, if $z, z' \in S(y)$ and $z \preceq y \preceq x$ so by hypothesis then $z \leq_{llex} z'$, as required. Secondly $P$ is linearly ordered. For otherwise if $x, a \in P$ with $x \| a$, then let $z$ be their $\prec$-maximal common ancestor. Consider two successors of $z$ say $v$ and $w$ with $v \prec x$ and $w \prec a$. Without loss of generality suppose that $v <_{llex} w$. Then $z$, $v$ and $w$ form a counterexample to $a$'s membership in $P$. Finally $P$ is infinite (and hence maximal with these properties). Indeed if $x \in P$, then the $<_{llex}$-smallest element in $S(x)$ is also in $P$. Hence $P$ is an infinite regular path in $\mathcal{T}$, as required. □

If in the hypothesis above $\mathcal{T}$ contains finitely many infinite paths, then every infinite path is regular since after defining $P$, one considers the tree on domain $T \setminus P$ to find the next infinite path. The next theorem generalises this to the case when $\mathcal{T}$ contains countably many infinite paths.

**Theorem 8.3 (Automatic König's Lemma, Version** 2) *If $\mathcal{T} = (T, \preceq)$ is a finitely branching automatic tree with countably many infinite paths, then every infinite path is regular.*

**Proof** Assume that $\mathcal{T}$ is pruned. Then the derivate $d(\mathcal{T})$ is definable and so the elements of the tree $\mathcal{T} \setminus d(\mathcal{T})$ form a regular subset of $T$, call it $R$. Then $R$ consists of countably many disjoint infinite paths, each definable as follows. For every $\prec$-minimal $a \in R$ define the infinite path $P_a$ as $\{x \in T \mid x \preceq a \vee (a \prec x \wedge x \in R)\}$.

Now replace $\mathcal{T}$ by $d(\mathcal{T})$ and repeat the steps in the previous paragraph. Since $CB(\mathcal{T})$ is finite, these steps can be iterated at most $CB(\mathcal{T})$ times; after which time the resulting tree will be empty and every infinite path in the original $\mathcal{T}$ will have been generated at some stage. □

**Remark 8.4** *The assumption that $\mathcal{T}$ have countably many infinite paths can not be dropped, since otherwise $\mathcal{T}$ necessarily has non-regular (indeed, uncountably many non-computable) infinite paths.*

## 8.2 The general case

It turns out that automaticity allows one to remove the condition that $\mathcal{T}$ be finitely branching, under the assumption of course that $\mathcal{T}$ has at least one infinite path. This can be done if given an automatic tree $\mathcal{T}$, one can effectively construct an automatic copy of the pruned tree $E(\mathcal{T})$, the set of elements of $\mathcal{T}$ that are on an infinite path in $\mathcal{T}$. Then as in the finitely branching case, Theorem 8.2, the $<_{llex}$-least path is definable and hence regular.

**Theorem 8.5 (Automatic König's Lemma, Version** 3) *If an automatic tree has an infinite path, then it has a regular infinite path.*

This follows immediately from the following construction.

**Lemma 8.6** *If $\mathcal{T}$ is an automatic tree then $E(\mathcal{T}) \subset T$ is a regular language.*

**Proof** Let $\mathcal{T} = (T, \preceq)$ be an automatic tree. Writing $T'$ for $E(\mathcal{T})$, it is required that the set $T' \subset \Sigma^\star$ of all nodes in $T$ that are on an infinite path is a regular language.

Recall that a (non-deterministic) Büchi automaton $(S, \iota, \Delta, F)$ over $\Sigma$ accepts an infinite string $\alpha \in \Sigma^\omega$ if it has a run $(q_i)_{i\in\mathbb{N}}$ such that there is some state $f \in F$ with $f = q_j$ for infinitely many $j \in \mathbb{N}$. The idea now is to construct a Büchi recognisable $\mathcal{B}$ over the alphabet $\Delta = \Sigma_\diamond \times \Sigma$ so that its projection (on the first co-ordinate) is of the form $T' \cdot \{\diamond\} \cdot W^\omega$ for some regular $W \subset \Sigma_\diamond^\star$. Then $T'$ is regular since Büchi automata are closed under projection and an automaton for $T'$ can be extracted from one for $B$.

Say that a word $x$ *is on* $c_0c_1\ldots$, where each $c_i$ is $(a_i, b_i) \in \Sigma_\diamond \times \Sigma$, iff there exist $m, n \in \mathbb{N}$ such that

- either $m = 0$, $x = a_0a_1\ldots a_n$ and $a_{n+1} = \diamond$
- or $n \geq m > 0$, $x = b_0b_1\ldots b_{m-1}a_ma_{m+1}\ldots a_n$, $a_{m-1} = \diamond$ and $a_{n+1} = \diamond$.

In the first case we say that $x$ is the *first word on* $c_0c_1\ldots$. Consider the set of all sequences $(a_0, b_0)(a_1, b_1)\ldots \in \Delta$ such that there are infinitely many words on the sequence and the words on the sequence generate an infinite path of $T$. More formally,

- $\exists^\infty n(a_n = \diamond)$;
- if $y, z$ are on $(a_0, b_0)\,(a_1, b_1)\,\ldots$ and $|y| \leq |z|$ then $y \preceq z$ and $y, z \in T$.

There is a Büchi automaton $\mathcal{B}$ accepting such sequences because the orderings $\preceq$ and length-comparison are automatic and $T$ is regular. Further using that $\mathcal{T}$ is transitive, one need only check that adjacent words $y, z$ on the sequence satisfy $y \preceq z$.

To complete the proof we prove that $x \in T'$ if and only if $x$ is the first word on some sequence $c_0c_1\ldots$ satisfying the two conditions. The reverse implication is clear. For the forward implication let $x \in T$ be given and $P$ be an infinite path witnessing that $x \in T'$. Define the sequences $a_0a_1\ldots$ and $b_0b_1\ldots$ described below.

1. **Choose** $n, a_0, a_1, \ldots, a_n$ such that $x = a_0a_1\ldots a_n$. **Let** $a_{n+1} = \diamond$.
2. **Let** $m = 0$. **Let** $y = x$.

3. **Find** $b_m b_{m+1} \dots b_{n+1}$ such that infinitely many nodes in $P$ extend $b_0 b_1 \dots b_{n+1}$ as strings.

4. **Update** $m = n + 2$.

5. **Find** a new value for $n$ and $a_m a_{m+1} \dots a_n$ such that $n \geq m$, the path $P$ contains the node $z = b_0 b_1 \dots b_{m-1} a_m a_{m+1} \dots a_n$ and $y \preceq z$. **Let** $a_{n+1} = \diamond$.

6. **Let** $y = z$. **Go to** 3.

Note that it is an invariant of the construction that whenever the algorithm comes to Step 3, either $m = 0$ or infinitely many nodes in $P$ extend the string $b_0 b_1 \dots b_{m-1}$. As there are only finitely many choices for the new part $b_m b_{m+1} \dots b_{n+1}$, one can choose this part such that still infinitely many nodes in $P$ extend $b_0 b_1 \dots b_{n+1}$ as a string. In Step 4, $m$ is chosen such that the precondition of Step 3 holds again and $b_m$ is the first of the $b$-symbols not yet defined. For every $y \in P$ it holds that all but finitely many nodes $z$ in $P$ satisfy $y \preceq z$. Furthermore, for every finite length $l$, almost all nodes in $P$ are represented by strings longer than $l$. Thus one can find a node $z$ as specified in Step 5 and the algorithm runs forever defining the infinite sequence $(a_0, b_0)(a_1, b_1) \dots$ in the limit. In particular, such a sequence exists. It is not required that the sequence can be constructed effectively since the path $P$ might not even be computable. □

From Theorem 8.5, we see that if an automatic tree has *finitely* many infinite paths, then each is regular. The next theorem generalises this to trees with *countably* many infinite paths.

**Theorem 8.7 (Automatic König's Lemma, Version** 4**)** *If an automatic tree has countably many infinite paths then every infinite path in it is regular.*

**Proof** Let $\mathcal{T} = (T, \preceq)$ be an automatic tree with countably many infinite paths. Then the extendible part of $\mathcal{T}$, $E(\mathcal{T}) \subset T$, is regular by Lemma 8.6. So the derivative $d(\mathcal{T})$ is automatic. Write $F^i(\mathcal{T})$ for the domain of $d^i(\mathcal{T})$ and $E^i(\mathcal{T}) \subset T$ for the extendible part of the domain of $d^i(\mathcal{T})$. Then since $\mathcal{T}$ is automatic $CB(\mathcal{T})$ is finite, say $n$. And since $\mathcal{T}$ has countably many infinite paths, $d^n(\mathcal{T})$ is the empty tree. So the structure $(T, F^1(\mathcal{T}), F^2(\mathcal{T}), \dots, F^n(\mathcal{T}), \preceq)$ is automatic.

Now for every $x \in \mathcal{T}$ there exists an $m < n$ such that $x \in E^m(\mathcal{T}) \setminus E^{m+1}(\mathcal{T})$. In particular if $P$ is an infinite path of $\mathcal{T}$ then there is a smallest $m \leq n$ such that $P \cap F^m(\mathcal{T})$ is infinite. In this case define $x_P$ to be the least, with respect to $\preceq$, element of $x \in F^m(\mathcal{T}) \setminus F^{m+1}(\mathcal{T})$ with $x_P \in P$. Then $P$ can be defined as the set of all $b \in F^m(\mathcal{T}) \setminus F^{m+1}(\mathcal{T})$ such that either $x_P \preceq b$ or $b \preceq x_P$. Hence $P$ is regular. □

Define a formula

$$\phi(a, b) = \bigvee_{i=0}^{n} (a \in F^i(\mathcal{T}) \wedge a \notin F^{i+1}(\mathcal{T}) \wedge b \in F^i(\mathcal{T}) \wedge (b \preceq a \vee a \preceq b)).$$

The formula tells for every infinite path $P$ and almost every $a \in P$ which nodes $b$ are on $P$. So $\{b \mid \phi(a, b)\}$ is either an infinite path of $\mathcal{T}$ or empty (in case that $a$ is not on an infinite path of $\mathcal{T}$). Thus one has the following result.

**Theorem 8.8** *If $\mathcal{T} = (T, \preceq)$ is an automatic tree with countably many infinite paths, then there is a formula $\phi$ such that the sets $P_a = \{b \in T \mid \Phi(a, b)\}$ satisfy the following conditions:*

- *If $P_a$ is not empty then $P_a$ is an infinite path of $(T, \preceq)$ containing $a$;*
- *Every infinite path of $(\mathcal{T}, \preceq)$ is equal to some $P_a$.*

**Acknowledgement.** The authors thank Wolfgang Merkle for useful discussions.

# References


[1] Achim Blumensath. *Automatic Structures.* Diploma thesis, RWTH Aachen, 1999.

[2] Achim Blumensath and Erich Grädel. Automatic structures. In *15th Symposium on Logic in Computer Science* (LICS 2000), pages 51–62, June 2000.

[3] Chirstian Delhommé. Non-automaticity of $\omega^\omega$, 2001. Manuscript.

[4] Christian Delhommé, Valentin Goranko and Teodor Knapik. Automatic linear orderings, 2003. Manuscript.

[5] Samuel Eilenberg, Calvin C. Elgot and John C. Shepherdson. Sets recognised by $n$–tape automata. *Journal of Algebra*, 13(4):447–464, 1969.

[6] David B. A. Epstein, James W. Cannon, Derek F. Holt, Silvio V. F. Levy, Michael S. Paterson and William P. Thurston. *Word processing in groups.* Jones and Bartlett, 1992.

[7] Hajime Ishihara, Bakhadyr Khoussainov and Sasha Rubin. Some results on automatic structures. In *17th Symposium on Logic in Computer Science* (LICS 2002), pages 235–242, July 2002.

[8] Alexander S. Kechris. *Classical Descriptive Set Theory.* Springer-Verlag, 1995.

[9] Bakhadyr Khoussainov and Anil Nerode. Automatic presentations of structures. In Daniel Leivant, editor, *Logic and Computational Complexity,* International Workshop LCC 1994, Indianapolis, IN, USA, volume 960 of *Lecture Notes in Computer Science*, pages 367–392, 1994.

[10] Bakhadyr Khoussainov, André Nies and Sasha Rubin. The isomorphism problem for automatic structures, 2003. Manuscript.

[11] Bakhadyr Khoussainov and Sasha Rubin. Graphs with automatic presentations over a unary alphabet. *Journal of Automata, Languages and Combinatorics*, 6(4), 2001.

[12] Bakhadyr Khoussainov and Sasha Rubin. Automatic structures – overview and future directions. *Journal of Automata, Languages and Combinatorics*, 8(2):287–301, 2003.

[13] Hartley Rogers. *Theory of Recursive Functions and Effective Computability.* McGraw-Hill, 1967.

[14] Joseph G. Rosenstein. *Linear Orderings.* Academic Press, 1982.