CDMTCS Research Report Series

Disjunctive Sequences: An Overview

Cristian S. Calude
Department of Computer Science
University of Auckland
Auckland, New Zealand

Lutz Priese
Universität Koblenz-Landau, Germany

Ludwig Staiger
Martin-Luther-Universität
Halle-Wittenberg, Germany

CDMTCS-063
October 1997

Centre for Discrete Mathematics and
Theoretical Computer Science

Disjunctive Sequences: An Overview*

Cristian S. Calude,† Lutz Priese,‡ Ludwig Staiger§

Abstract

Following Jürgensen and Thierrin [21] we say that an infinite sequence is *disjunctive* if it contains any (finite) word, or, equivalently, if any word appears in the sequence infinitely many times. "Disjunctivity" is a natural qualitative property; it is weaker, than the property of "normality" (introduced by Borel [1]; see, for instance, Kuipers, Niederreiter [24]). The aim of this paper is to survey some basic results on disjunctive sequences and to explore their role in various areas of mathematics (e.g. in automata-theoretic studies of ω-languages or number theory). To achieve our goal we will use various instruments borrowed from topology, measure-theory, probability theory, number theory, automata and formal languages.

1 Notation and Definitions

Let $\mathbb{N}$ be the set of positive integers. The number of elements of a finite set S is denoted by $\mathrm{card}(S)$. For any finite set (alphabet) X let X^* denote the free monoid of words (including the empty word ϵ) over X, and X^ω the set of (infinite) sequences over X. Words on X are denoted by u, v, w; sequences over X are denoted by $\mathbf{x}, \mathbf{y}, \mathbf{z}$. For $W \subseteq X^*$ the submonoid generated by W is denoted W^*, and W^ω is the set of infinite sequences formed by concatenating members of W; finally, let $X^\infty = X^* \cup X^\omega$. A subset $W \subseteq X^*$ is called a language; an ω-language is a subset of X^ω. For $w \in X^*$ and $\gamma \in X^\infty$ the concatenation of w and γ is written $w\gamma$. This defines in an obvious way a product $W\Gamma$ of sets $W \subseteq X^*$ and $\Gamma \subseteq X^\infty$: $W\Gamma = \{w\gamma \mid w \in W, \gamma \in \Gamma\}$. For the sake of brevity we shall write wB, w^* and w^ω instead of $\{w\}B$, $\{w\}^*$ and $\{w\}^\omega$, respectively. By $|w|$ we denote the length of the word $w \in X^*$. The set of all *initial words (prefixes)* of $\gamma \in X^\infty$ is

$$\mathbf{A}(\gamma) = \{w \in X^* \mid \exists \gamma' \in X^\infty \;\; w\gamma' = \gamma\}.$$

*The first author has has been partially supported by Auckland University Research Grant, A18/XXXXX/62090/F3414050.

†Computer Science Department, The University of Auckland, Private Bag 92109, Auckland, New Zealand; email: cristian@cs.auckland.ac.nz.

‡Universität Koblenz-Landau, Fachbereich Informatik, Rheinau 1, D-56075 Koblenz, Germany, email: priese@informatik.uni-koblenz.de.

§Martin-Luther-Universität Halle-Wittenberg, Institut für Informatik, Kurt-Mothes-Str. 1, D-06120 Halle (Saale), Germany, email: staiger@cantor.informatik.uni-halle.de.

The set of all *subwords (factors, infixes)* of $\gamma \in X^\infty$ is

$$\mathbf{T}(\gamma) = \{w \in X^* \mid \exists v \in X^* \ \exists \gamma' \in X^\infty \ \ vw\gamma' = \gamma\},$$

and the set of all *suffixes* of $\gamma \in X^\infty$ is

$$\mathbf{S}(\gamma) = \{\gamma' \in X^\infty \mid \exists w \in X^*, \ \ w\gamma' = \gamma\}.$$

For $B \subseteq X^\infty$ put

$$\mathbf{A}(B) = \bigcup_{\gamma \in B} \mathbf{A}(\gamma), \quad \mathbf{T}(B) = \bigcup_{\gamma \in B} \mathbf{T}(\gamma), \quad \mathbf{S}(B) = \bigcup_{\gamma \in B} \mathbf{S}(\gamma),$$

and let $w \sqsubseteq \gamma$ denote $w \in \mathbf{A}(\gamma)$. For $W \subseteq X^*$ the *δ-limit* of W, W^δ, consists of all infinite sequences of X^ω that contain infinitely many prefixes in W,

$$W^\delta = \{\mathbf{x} \in X^\omega \mid \mathbf{A}(\mathbf{x}) \cap W \text{ is infinite}\}.$$

The *adherence* of $B \subseteq X^\infty$ is defined by

$$\operatorname{adh} B = (\mathbf{A}(B))^\delta.$$

Obviously, $\operatorname{adh} B = \{\mathbf{x} \in X^\omega \mid \mathbf{A}(\mathbf{x}) \subseteq \mathbf{A}(B)\}$ holds true.

For $\gamma \in X^\infty$ we denote by $\gamma[k]$ the prefix of γ of length k and by $\gamma[k, \ell]$ the infix from the k-th to the ℓ-th letter of γ, i.e., $\gamma = \gamma[k-1]\gamma'$ implies $\gamma[k, \ell] = \gamma'[\ell - k + 1]$; $\gamma[0] = \epsilon$.

For $B \subseteq X^\infty$ we define the *state* B/w of B generated by the word $w \in X^*$ as $B/w = \{b \mid wb \in B\}$. A set B is called *finite-state* if its set of states $\{B/w \mid w \in X^*\}$ is finite.

A finite-state language $W \subseteq X^*$ is also called *regular.*[1] An ω-language F is called *regular* provided there is an $n \in \mathbb{N}$ and $2n$ regular languages W_i, V_i $(1 \le i \le n)$ such that

$$F = \bigcup_{i=1}^{n} W_i V_i^\omega. \tag{1}$$

Similarly, an ω-language F is called *context-free* provided there is an $n \in \mathbb{N}$ and $2n$ context-free languages W_i, V_i $(1 \le i \le n)$ such that the equation (1) holds true.

Since a nonempty ω-language of the form W^ω is either $W^\omega = \{w\}^\omega = \{w^\omega\}$, for some $w \in W$, or, otherwise, is of the cardinality of the continuum, every at most countable regular or context-free ω-language only consists of ultimately periodic sequences.

The *Baire-metric* ρ on X^∞ is defined by

$$\rho(\gamma, \gamma') = \inf \{(\operatorname{card} X)^{-|w|} \mid w \in \mathbf{A}(\gamma) \cap \mathbf{A}(\gamma')\}.$$

Notice that this metric ρ satisfies the ultra-metric inequality $\rho(\gamma, \gamma') \le \max\{\rho(\gamma, \gamma''), \rho(\gamma', \gamma'')\}$; it defines a topology on X^∞ where the open sets are of the form $UX^\omega \cup V$ for $U, V \subseteq X^*$ and the closed sets are of the form $\operatorname{adh} U \cup V$ for $U, V \subseteq X^*$. However, in this paper we deal mainly with X^ω. We consider X^ω as a topological (ultra-metric) space with the topology induced from X^∞. Thus, open sets in X^ω are of the form UX^ω and closed sets of the form $\operatorname{adh} U$, where $U \subseteq X^*$. In addition, a set E in X^ω is closed if and

[1] In fact, regularity of $W \subset X^*$ is usually defined in a different way, but it is well known that a language W is regular if and only if it is finite state.

only if $E = \mathrm{adh}\, U$ for some $U \subseteq X^*$, if and only if $E = \mathrm{adh}\, E$. E is *clopen* in X^ω, that is, simultaneously open and closed, if and only if $E = UX^\omega$ for some finite set $U \subseteq X^*$. The open balls in (X^ω, ρ) are the sets of the form wX^ω, $w \in X^*$. They are simultaneously closed. The ball wX^ω has diameter $\mathrm{diam}(wX^\omega) = (\mathrm{card}\, X)^{-|w|}$, hence X^ω is a Cantor space induced by the basis $\{wX^\omega \mid w \in X^*\}$. Since X is finite, this topological space is homeomorphic to the Cantor discontinuum, hence compact. The topological closure $\mathcal{C}(K)$ of any subset $K \subseteq M$ of some topological space M is the smallest closed subset of M containing K. Thus, $\mathcal{C}(K) = \mathrm{adh}\, K$ holds in X^ω for $K \subseteq X^\omega$.

Having defined open and closed sets in X^ω, we proceed to the next classes of Borel's hierarchy (cf. Kuratowski [25]): $\mathbf{F}$ is the set of closed subsets of X^ω, $\mathbf{G}$ is the set of open subsets of X^ω, $\mathbf{F}_\sigma$ is the set of countable unions of closed subsets of X^ω, $\mathbf{G}_\delta$ is the set of countable intersections of open subsets of X^ω, $\mathbf{F}_{\sigma\delta}$ is the set of countable intersections of sets in $\mathbf{F}_\sigma$, and $\mathbf{G}_{\delta\sigma}$ is the set of countable unions of sets in $\mathbf{G}_\delta$.

A sequence $\mathbf{x} \in X^\omega$ is called *disjunctive* if every word in X^* appears as a subword in $\mathbf{x}$, that is, $\mathbf{T}(\mathbf{x}) = X^*$. Equivalently, $\mathbf{x} \in X^\omega$ is disjunctive if every word in X^* appears infinitely often as a subword in $\mathbf{x}$.

Let $b \geq 2$ be an integer and put $X_b = \{0, 1, \ldots, b-1\}$. To any sequence $x_0 x_1 \ldots x_n \ldots \in X_b^\omega$ one can associate the real number

$$v_{\mathbf{x}} = \sum_{i=0}^{\infty} x_i b^{-i-1}.$$

Conversely, the base–b expansion $s_b(x)$ of the real number x in the interval $[0,1)$ is the unique infinite sequence $x_0 x_1 \ldots x_n \ldots \in X_b^\omega$ containing infinitely many digits different from $b-1$ such that $v_{\mathbf{x}} = x$. With the above choice of the metric ϱ in the Cantor space (X_b^ω, ϱ) we have $|v_{\mathbf{x}} - v_{\mathbf{y}}| \leq \varrho(\mathbf{x}, \mathbf{y})$. Let $I_{b,w} \subseteq [0,1]$ denote the largest open interval contained in the real image of the ball wX_b^ω. Then $I_{b,w} = (v_{\mathbf{x}}, v_{\mathbf{y}})$ where $\mathbf{x} = w0^\omega$ and $\mathbf{y} = w(b-1)^\omega$. Moreover, $F \cap wX_b^\omega = \emptyset$ implies that the real image of $F \subseteq X_b^\omega$, $\{v_{\mathbf{x}} \mid \mathbf{x} \in F\}$, is disjoint from $I_{b,w}$.

Finally, a real number $x \in [0,1)$ is *disjunctive to base b* if $s_b(x) \in X_b^\omega$ is disjunctive; x is *absolutely disjunctive* in case it is disjunctive to any base.

2 Examples

Consider first the alphabet X_b. Champernowne's sequence [12] over X is defined by concatenating all words over X_b in some recursive order, say in the quasi-lexicographical one:[2]

$$01 \cdots (b-1)0001 \cdots 0(b-1)1011 \cdots (b-1)(b-1)000001 \cdots \tag{2}$$

Champernowne proved that for $b = 10$ the above sequence is normal[3] in the scale of ten, so it is disjunctive. In fact, (2) is normal in every base b^n, $n \geq 2$. *It is not known whether this sequence is normal in any scale from b^n, $n \geq 2$.*

A related example is due to Copeland and Erdös [13]: the sequence of primes

$$23571113171923 \cdots$$

[2] Words are arranged in increasing order of their length; words having the same length are arranged lexicographically.

[3] Normality was introduced by Borel [1]; see also Kuipers, Niederreiter [24]).

is normal in the scale of ten, so it is disjunctive.

Not every disjunctive sequence is normal: for example, take the sequence

$$0w_0 00w_1 000w_2 0000 \cdots w_n 0^n \cdots,$$

where $w_0 w_1 w_2 \cdots w_n \cdots$ is Champernowne's sequence over the alphabet X_b and 0^n abbreviates the word of length n containing only zeros.

The following result, due to Istrate and Păun [20], can be used to generate many disjunctive sequences.

Lemma 2.1. *Let $(a_n)_n$ be a strictly increasing sequence of positive integers such that*

$$\lim_{n\to\infty} \frac{a_{n+1}}{a_n} = 1, \tag{3}$$

and let $b \geq 2$ be a base. If we denote by b_n the base–b expansion of a_n, then the infinite sequence obtained by concatenating $b_1, b_2, \ldots, b_n, \ldots$ (in this order) is disjunctive over X_b.

Proof. Pick a word $w \in X_b^*$ and assume, without loss of generality, that w does not start with 0. Let w be the base–b expansion of m. In view of (3) there exists an integer $n_w \geq 1$ such that for all $n \geq n_w$,

$$\frac{a_{n+1}}{a_n} < \frac{m+1}{m}.$$

For every positive integer k define $i_k = \max\{n \mid a_n < mb^k\} \cup \{0\}$ and notice that for all k,

$$a_{i_k} < mb^k \leq a_{i_k+1}.$$

Furthermore, $\lim_{k\to\infty} i_k = \infty$. For $i_k > n_w$, $ma_{i_k+1} < a_{i_k}(m+1)$, so

$$mb^k \leq a_{i_k+1} < a_{i_k}(m+1)/m < (m+1)b^k,$$

hence

$$mb^k \leq a_{i_k+1} < (m+1)b^k.$$

To conclude, the base–b expansion of $a_{i_{k+1}}$ starts with w, so the sequence is disjunctive over X. □

As an example take $a_n = n$, for all n, and $b = 2$. The resulting sequence is

$$0110111001011101111000\cdots$$

If $w_0 w_1 w_2 \cdots w_n \cdots$ is Champernowne's binary sequence $0100011011000001\cdots$, then the above sequence is exactly $11w_0 1w_1 1w_2 \cdots$.

Consider now an infinite sequence

$$\mathbf{x} = x_0 x_1 \cdots \in \{0,1\}^\omega$$

and, following von Mises [51], define a new infinite sequence

$$\mathbf{y} = y_0 y_1 \cdots \in \{0,1,2\}^\omega$$

by the formula

$$y_i = \begin{cases} x_0, & \text{if } i = 0, \\ x_{i-1} + x_i, & \text{if } i > 1. \end{cases}$$

The sequence $\mathbf{y}$ is not disjunctive, even if $\mathbf{x}$ was chosen disjunctive: an infinity of words do not appear in $\mathbf{y}$, for example, the words 02 or 20. A seemingly minor change in the above example makes a major change;[4] define:

$$\mathbf{z} = z_0 z_1 \cdots \in \{0,1\}^{\omega}$$

by

$$z_i = \begin{cases} x_0, & \text{if } i = 0, \\ x_{i-1} \oplus x_i, & \text{if } i \geq 1. \end{cases}$$

If $\mathbf{x}$ is disjunctive, then so is $\mathbf{z}$. Indeed, the words 00, 01, 10, and 11 are transformed into 0, 1, 1, 0, respectively. Assume that x is transformed into y. If x ends with 0, then $x0$ transforms to $y0$, and $x1$ reduces to $y1$; if x ends in 1, then $x0$ goes in $y1$ and $x1$ goes in $y0$. A proof by induction shows that $\mathbf{z}$ is disjunctive provided $\mathbf{x}$ is disjunctive.

We finish this section with a uniform way to construct disjunctive sequences. Consider a recursive bijection $S : \mathbb{N} \to X^*$ and surjective function $f : \mathbb{N} \to \mathbb{N}$. The sequence

$$S(f(0))S(f(1)) \cdots S(f(n)) \cdots \tag{4}$$

is disjunctive. Conversely, if $\mathbf{x} = x_0 x_1 \cdots x_n \cdots$ is disjunctive, then we can find a strictly increasing sequence of non-negative integers $(t_i)_{i \geq 0}$ such that $\mathbf{x} = S(t_0)S(0)S(t_1)S(1) \cdots S(t_n)S(n) \cdots$, so the formula (4) works for the function $f(0) = t_0, f(1) = 0, f(2) = t_1, f(3) = 1, \ldots$.

3 The ω-Language of Disjunctive Sequences

In this section we will present a few simple properties of the ω-language of all disjunctive sequences over X, $D = \{\mathbf{x} \mid \mathbf{T}(\mathbf{x}) = X^*\}$.

3.1 Basic properties

From the very definition of disjunctive sequences we obtain

$$D = \bigcap_{w \in X^*} X^* w X^{\omega}. \tag{5}$$

Lemma 3.1. *The ω-language D is finite-state but not regular.*

Proof. Since $w\mathbf{x} \in D$ if and only if $\mathbf{x} \in D$, the ω-language D satisfies $D/w = D$, for all $w \in X^*$. Thus D has only a single state. Next, D is nonempty and does not contain an ultimately periodic sequence wv^{ω}. Following (1) the ω-language D cannot be regular. □

[4]This transformation was suggested by A. Szilard; here $\oplus$ stands for the sum modulo 2. Szilard's transformation also preserves random sequences, cf. Calude and Jürgensen [7].

We turn our attention to recursion theoretic properties of D. To this end we introduce the first classes of the arithmetical hierarchy of ω-languages. As usual we say that an ω-language $E \subseteq X^\omega$ is Π_1-*definable* provided E is representable in the form

$$E = \{\mathbf{x} \in X^\omega \,|\, \forall w (w \sqsubseteq \mathbf{x} \Rightarrow w \in W_E)\}, \tag{6}$$

where $W_E \subseteq X^*$ is a recursive language, and we say that an ω-language $F \subseteq X^\omega$ is Π_2-*definable* provided F is representable in the form

$$F = \{\mathbf{x} \in X^\omega \,|\, \forall w (w \in X^* \Rightarrow \exists u (u \sqsubseteq \mathbf{x} \wedge (w, u) \in M_F))\}, \tag{7}$$

where M_F is a recursive subset of $X^* \times X^*$.

It is well-known that in Cantor's topology, Π_1-definable ω-languages are closed sets and Π_2-definable ω-languages are $\mathbf{G}_\delta$-sets.

Lemma 3.2. *The ω-language of all disjunctive sequences D is Π_2-definable.*

Proof. We have $D = \{\mathbf{x} \in X^\omega \,|\, \forall w \exists v (vw \sqsubseteq \mathbf{x})\}$. So it suffices to put $M_D = \{(w, vw) \,|\, w, v \in X^*\}$ in (7). □

Thus we have seen that D is a $\mathbf{G}_\delta$-set. For $\mathbf{G}_\delta$-sets we have the following characterization via languages (cf. Thomas [50]).

Theorem 3.3. *In Cantor's topology, a subset $F \subseteq X^\omega$ is a $\mathbf{G}_\delta$-set if and only if there is a language $W \subseteq X^*$ such that $F = W^\delta$.*

The preceding theorem explains also why we called W^δ the δ-limit of the language W.

In Staiger [44], Proposition 7.6, it is shown that *an ω-language $F \subseteq X^\omega$ is Π_2-definable if and only if there is a recursive language $W \subseteq X^*$ such that $F = W^\delta$*. In case of D we can construct W_D explicitly.

Proposition 3.4. *Let*

$$W_D = \{wx \,|\, w \in X^* \wedge x \in X \wedge \exists n (n \leq |w| + 1 \wedge \mathbf{T}(wx) \supseteq X^n \wedge \mathbf{T}(w) \not\supseteq X^n)\}.$$

Then W_D is a recursive language and $D = W_D^\delta$.

Proof. It is obvious that W_D is recursive. Let $\mathbf{x}$ be a sequence such that $\mathbf{T}(\mathbf{x}) = X^*$. Then for every $n \geq 1$ there is a shortest prefix $w_n \sqsubseteq \mathbf{x}$ such that $\mathbf{T}(w_n) \supseteq X^n$. Thus $\{w_n \,|\, n \geq 1\}$ is an infinite subset of W_D. The converse implication follows from the observation that if $u, v \in W_D$ and $u \sqsubset v$, then $X^m \subseteq \mathbf{T}(u)$ implies $X^m \subseteq \mathbf{T}(v)$, and there is an $n \in \mathbb{N}$ satisfying $\mathbf{T}(u) \not\supseteq X^n \subseteq \mathbf{T}(v)$. □

We conclude this part by showing that in Cantor's topology D is not an $\mathbf{F}_\sigma$-set. To this end we quote Theorem 21 from Staiger [42].

Theorem 3.5. *If $F \subseteq X^\omega$ is finite-state and simultaneously an $\mathbf{F}_\sigma$- and a $\mathbf{G}_\delta$-set, then F is regular.*

Combining Theorem 3.5 with Lemmas 3.1 and 3.2 we get:

Proposition 3.6. *In Cantor's topology, D is not an $\mathbf{F}_\sigma$–set.*

We continue by deriving two topological characterizations of D.

3.2 A metric related to languages

The first topology is related to the well-known fact that every $\mathbf{G}_\delta$-set of a complete metric space is a complete metric space itself (cf. Kuratowski [25]). We use here the construction presented in Staiger [45].

As we have seen in Theorem 3.3, in Cantor's topology a $\mathbf{G}_\delta$-set is of the form U^δ for some $U \subseteq X^*$. We use this language U to define a new metric ϱ_U on X^ω which makes U^δ a closed set in the metric space (X^ω, ϱ_U):

$$\varrho_U(\mathbf{x}, \mathbf{y}) = \begin{cases} 0 & \text{, if } \mathbf{x} = \mathbf{y}, \\ (\operatorname{card} X)^{1-\operatorname{card} \mathbf{A}(\mathbf{x}) \cap \mathbf{A}(\mathbf{y}) \cap U} & \text{, otherwise.} \end{cases} \tag{8}$$

This metric, in some sense, resembles the metric ϱ of the Cantor's space; in fact, $\varrho = \varrho_{X^*}$. In analogy with $\mathcal{C}(F)$ we denote by $\mathcal{C}_U(F)$ the smallest closed (with respect to ϱ_U) subset of X^ω containing F. A point $\mathbf{x} \in \mathcal{C}_U(F)$ is called an *isolated point* of F provided $\exists \varepsilon(\varepsilon > 0 \wedge \forall \mathbf{y}(\mathbf{y} \in F \wedge \mathbf{y} \neq \mathbf{x} \Rightarrow \varrho_U(\mathbf{x}, \mathbf{y}) > \varepsilon))$. A point $\mathbf{x} \in \mathcal{C}_U(F)$ which is not isolated is called a *cluster point* of F.[5]

Theorem 3.7. *Let $U \subseteq X^*$. Then U^δ is closed in (X^ω, ϱ_U).*

Proof. It suffices to show that every point $\mathbf{x} \notin U^\delta$ is isolated. Let $\operatorname{card} \mathbf{A}(\mathbf{x}) \cap U = n < \infty$. From the definition of ϱ_U we have $\varrho_U(\mathbf{x}, \mathbf{y}) \geq (\operatorname{card} X)^{1-n}$ for all $\mathbf{y} \in X^\omega$. □

In case of D we can prove even more:

Theorem 3.8. *In the space $(X^\omega, \varrho_{W_D})$ the ω-language D equals both the set of its cluster points and the set of cluster points of the whole space X^ω.*

Proof. From the preceding proof we know that every point of $X^\omega \setminus D$ is an isolated point in $(X^\omega, \varrho_{W_D})$. Thus in view of Theorem 3.7 it remains to show that no point of $D = W_D^\delta$ is isolated. Let $\mathbf{x} \in D$. Then for every $w \sqsubseteq \mathbf{x}$ we have also $w\mathbf{x} \in D$. Thus $(w\mathbf{x})_{w \in \mathbf{A}(\mathbf{x}) \cap W_D}$ is an infinite subfamily of D such that $\inf\{\varrho_{W_D}(\mathbf{x}, w\mathbf{x}) \mid w \in \mathbf{A}(\mathbf{x}) \cap W_D\} = 0$. □

3.3 The topology of forbidden words

Next we introduce the second topology $\mathcal{T}$ related to disjunctive sequences. Since this topology is not a metric one, we introduce it by its set of open sets.

Let $\mathcal{O}_\mathcal{T} = \{X^* W X^\omega \mid W \subseteq X^*\}$. This family $\mathcal{O}_\mathcal{T}$ is closed under finite intersections and arbitrary unions, thus it may be considered as the set of open sets for a topology on X^ω.

An ω-language $F \subseteq X^\omega$ *avoids* words of a language $W \subseteq X^*$ provided $F \subseteq X^\omega \setminus X^* W X^\omega$, that is, no word $w \in W$ occurs as a subword (infix) of an ω-word $\mathbf{x} \in F$. Therefore, the complements of open sets in our topology $\mathcal{T}$ are characterized by the fact that their ω-words do not contain subwords from W. Thus $\mathcal{T}$ will be called the

[5] It should be mentioned that every set of isolated points in a metric space is open.

topology of “forbidden” words. In particular, sets open in this topology are also open in the Cantor’s topology of X^ω.[6]

Sets closed in the topology of forbidden words can be characterized in a similar way as closed sets in Cantor’s topology.

Theorem 3.9. *Let $F \subseteq X^\omega$. Then the following conditions are equivalent:*

1. *F is closed in the topology of forbidden words.*
2. *F is closed in Cantor’s topology and $\forall w(w \in X^* \Rightarrow F \supseteq F/w)$.*
3. *F is closed in Cantor’s topology and $\mathbf{A}(F) = \mathbf{T}(F)$.*
4. *$\forall \mathbf{x}(\mathbf{A}(\mathbf{x}) \subseteq \mathbf{T}(F) \Rightarrow \mathbf{x} \in F)$.*

Proof. “*1.* $\Rightarrow$ *2*”: As we noticed above, every ω-language closed in the topology of forbidden words is also closed in Cantor’s topology. Let $w \in X^*$ and $F = X^\omega \setminus X^*WX^\omega$. Then $F/w = X^\omega \setminus (X^*WX^\omega)/w$, and the assertion follows from the obvious inclusion $(X^*WX^\omega)/w \supseteq X^*WX^\omega$.

“*2.* $\Rightarrow$ *3.*” follows from the identity $\mathbf{A}(\bigcup_{w\in X^*} F/w) = \mathbf{T}(F)$.

The implication “*3.* $\Rightarrow$ *4.*” is obvious.

Finally, we show that Condition 4 implies $F = X^\omega \setminus X^* \cdot (X^* \setminus \mathbf{T}(F)) \cdot X^\omega$. Since $X^* \setminus \mathbf{T}(F) = X^* \cdot (X^* \setminus \mathbf{T}(F)) \cdot X^*$ it suffices to prove that $F = X^\omega \setminus (X^* \setminus \mathbf{T}(F)) \cdot X^\omega$.

The inclusion $F \subseteq X^\omega \setminus (X^* \setminus \mathbf{A}(F)) \cdot X^\omega \subseteq X^\omega \setminus (X^* \setminus \mathbf{T}(F)) \cdot X^\omega$ follows from $\mathbf{A}(F) \subseteq \mathbf{T}(F)$. To prove the converse inclusion let $\mathbf{x} \notin F$ Then in view of Condition 4 there is a prefix $w \sqsubset \mathbf{x}$ such that $w \notin \mathbf{T}(F)$. Consequently, $\mathbf{x} \in (X^* \setminus \mathbf{T}(F)) \cdot X^\omega$. □

The additional requirements in conditions 2 and 3 are, however, not equivalent in general. The following example shows that there is an ω-language (necessarily not closed in Cantor’s topology) which satisfies $\mathbf{A}(F) = \mathbf{T}(F)$, but not the condition $\forall w(w \in X^* \Rightarrow F \supseteq F/w)$.

Example 3.10. *Let $F = (X^2)^*bba^\omega \cup X(X^2)^*aab^\omega$. Then $\mathbf{A}(F) = \mathbf{T}(F) = X^*$, but $F/a \not\subseteq F$.*

If the language of forbidden patterns $W \subseteq X^*$ is regular, then the ω-language $F_W = X^\omega \setminus X^*WX^\omega$ is a regular ω-language. In connection with (1) this yields as a consequence the following generalization of a result of El-Zanati and Transue [16].

Theorem 3.11. *Let $W \subseteq X^*$ be a regular language. If F_W is uncountable, then F_W contains a set homeomorphic to $\{a,b\}^\omega$, more precisely, a subset of the form $w\{u,v\}^\omega$, where $u \neq v$ and $|u| = |v| > 0$.*

We continue with some more examples. The first is an example of a countable regular ω-language F_W which requires an infinite set of forbidden patterns.

[6] Recall that open sets in Cantor’s topology are ω-languages of the form WX^ω, where $W \subseteq X^*$, and closed sets are characterized by the their initial word languages, that is, $F \subseteq X^\omega$ is closed if and only if $\mathbf{A}(\mathbf{x}) \subseteq \mathbf{A}(F)$ implies $\mathbf{x} \in F$.

Example 3.12. *Let $X = \{a, b\}$ and $W = ba^*b$. Then $F_W = X^\omega \setminus X^*WX^\omega = a^*ba^\omega \cup a^\omega$ is a countable ω-language. It is clear that $F_W \neq F_V$, for any finite language $V \subseteq X^*$.*

Though the regularity of W implies the regularity of F_W this same relation is not true for context-free languages and ω-languages.

Example 3.13. *Let $X = \{a, b\}$ and $W = \{bb\} \cup \{ba^iba^jb \mid j \neq i+1\}$. Clearly, W is a deterministic context-free language, and $F_W = a^*(\{\eta_i \mid i \in \mathbb{N}\} \cup \{\eta_{i,j} \mid i, j \in \mathbb{N} \wedge i \leq j\})$ where $\eta_i = ba^iba^{i+1}b\cdots$ and $\eta_{i,j} = ba^iba^{i+1}\cdots ba^jba^\omega$. Since F_W is countable but does not consist entirely of ultimately periodic ω-words, the equation* (1) *shows that F_W is not context-free.*

Finally, we discuss characterizations of disjunctive sequences and the ω-language D by means of the topology of forbidden words. From (5) we obtain immediately:

Proposition 3.14. *In the topology of forbidden words, D is the smallest nonempty $\mathbf{G}_\delta$-set.*

Proposition 3.15. *A sequence $\mathbf{x} \in X^\omega$ is disjunctive if and only if the set $\{\mathbf{x}\}$ is dense in X^ω in the topology of forbidden words.*

Proof. A set $F \subseteq X^\omega$ is dense in X^ω in case X^ω is the smallest closed set containing F, that is, $X^\omega \setminus F$ does not contain a nonempty open set. Since $\mathbf{x} \in X^\omega$ is disjunctive, we have $\mathbf{T}(\mathbf{x}) = X^*$, and therefore $\{\mathbf{x}\} \cap X^*wX^\omega \neq \emptyset$ for all $w \in X^*$. □

Above we mentioned, the topology of forbidden words is not a metric topology. Proposition 3.15 gives evidence of this fact, because in any metric topology every finite set is closed.

4 How Large is the ω-language of Disjunctive Sequences?

In this section we study the size of our set D. Appropriate size measures for sets in Cantor's space are topological density (Baire category) and (product) measure. It turns out that in both cases the ω-language of disjunctive sequences is a large set. Its complement $X^\omega \setminus D$ can be characterized using regular ω-languages. Therefore, we consider category and measure especially for regular and finite-state ω-languages. Moreover, for these ω-languages, density and the appearance of subwords are closely related.

4.1 Density and Baire category

A set F is *nowhere dense* in $E \subseteq X^\omega$ provided $\mathcal{C}(E \setminus \mathcal{C}(F)) = \mathcal{C}(E)$, that is, if $\mathcal{C}(F)$ does not contain a nonempty subset of the form $E \cap wX^\omega$. This condition can be reformulated as follows.

Lemma 4.1. *A set $F \subseteq X^\omega$ is nowhere dense in E if and only if for every $v \in \mathbf{A}(E)$ there is a $w \in X^*$ such that $vw \in \mathbf{A}(E)$ and $vwX^\omega \cap F = \emptyset$.*

Cast in the language of prefixes, Lemma 4.1 asserts that F is not nowhere dense in $E \neq \emptyset$ if and only if there is a $w \in \mathbf{A}(E)$ such that $E/w \subseteq \mathcal{C}(F)/w$. From the following equations

$$\mathcal{C}(E \setminus \mathcal{C}(F)) = \mathcal{C}(\mathcal{C}(E) \setminus \mathcal{C}(F)) \tag{9}$$

$$\mathcal{C}(E \setminus \mathcal{C}(F)) = \mathcal{C}(E \setminus (\mathcal{C}(F) \cap E)) \tag{10}$$

we see that F is nowhere dense in E if and only if F is nowhere dense in $\mathcal{C}(E)$ and if and only if $\mathcal{C}(F) \cap E$ is nowhere dense in E.

A subset $F \subseteq X^\omega$ is called nowhere dense if it is nowhere dense in X^ω. As usual we call a set F of *first Baire category (in E)* if it is a countable union of sets nowhere dense (in E).

In this section we consider the density of finite-state ω-languages $F \subseteq X^\omega$ in so-called ω-power languages W^ω.

We obtain the following version of Lemma 4.1.

Corollary 4.2. *Let $W \subseteq X^*$ be a nonempty language. Then F is nowhere dense in W^ω if and only if for every $v \in W^*$ there is a $w \in W^*$ such that $vwX^\omega \cap F = \emptyset$.*

Using again the language of prefixes, F is not nowhere dense in an ω-power W^ω if and only if there is a $w \in W^*$ such that $W^\omega/w \subseteq \mathcal{C}(F)/w$.

It turns out that for finite-state ω-languages density in ω-powers is closely related to the appearence of subwords. To this end we present some results of Staiger [47].

Lemma 4.3. *Let F be finite-state and let $W \subseteq X^*$ be nonempty. Then the following conditions are equivalent:*

1. *F is nowhere dense in W^ω.*
2. *$\forall w(w \in W^* \Rightarrow F/w$ is nowhere dense in $W^\omega)$.*
3. *$\forall v(v \in X^* \Rightarrow (\mathcal{C}(F) \cap \mathcal{C}(W^\omega))/v$ is nowhere dense in $W^\omega)$.*

Proof. The implication *2.* $\rightarrow$ *1.* is obvious.

Since $v \in W^*$ implies $\mathcal{C}(W^\omega) \subseteq \mathcal{C}(W^\omega)/v$, the implication *3.* $\rightarrow$ *2.* follows.

In order to prove the remaining *1.* $\rightarrow$ *3.* assume $(\mathcal{C}(F) \cap \mathcal{C}(W^\omega))/v$ to be not nowhere dense in W^ω. Then there is a $w \in W^*$ such that $(\mathcal{C}(F) \cap \mathcal{C}(W^\omega))/v \cdot w \supseteq W^\omega/w \supseteq W^\omega$. Define $u := v \cdot w$. Thus $\mathcal{C}(W^\omega)/u \supseteq \mathcal{C}(W^\omega)$ and $\mathcal{C}(F)/u \supseteq \mathcal{C}(W^\omega)$

Since F is finite-state, there are $n, k \geq 1$ such that $F/u^n = F/u^{n+k}$. Hence $\mathcal{C}(W^\omega) \subseteq \mathcal{C}(F)/u$ implies $\mathcal{C}(W^\omega)/u^{n+k-1} \subseteq \mathcal{C}(F)/u^{n+k} = \mathcal{C}(F)/u^n$. Finally observe that $\mathcal{C}(W^\omega)/u^n \subseteq \mathcal{C}(W^\omega)/u^{n+k-1} \subseteq \mathcal{C}(F)/u^n$, what proves our assertion. □

As a consequence of Lemma 4.3 we show that finite-like-state ω-languages which are nowhere dense in ω-power sets have patterns, that is subwords appearing in the ω-power set W^ω and do not appear in the finite-state ω-language F.

Lemma 4.4. *Let F be finite-state, and let $W \subseteq X^*$ be a nonempty language.*

1. *F is nowhere dense in W^ω if and only if there is a $w \in W^*$ such that $\mathcal{C}(F)\cap\mathcal{C}(W^\omega) \subseteq \mathcal{C}(W^\omega) \setminus W^* w X^\omega$.*
2. *If $F \subseteq \mathcal{C}(W^\omega)$ then F is nowhere dense in W^ω if and only if there is a $u \in W^*$ such that $F \subseteq \mathcal{C}(W^\omega) \setminus X^* u X^\omega$.*

Proof. *1.* If F is finite-state and nowhere dense in W^ω then according to Lemma 4.3 the set $F' = \bigcup_{u\in W^*} F/u$ as a finite union of sets nowhere dense in W^ω is again nowhere dense in W^ω. Hence, there is a $w \in W^*$ such that $F' \cap wX^\omega = \emptyset$. Assume now that $F \cap W^* w X^\omega \neq \emptyset$. Then there is some $v \in W^*$ such that $F \cap vwX^\omega = v(F/v) \cap vwX^\omega \neq \emptyset$, which contradicts the fact that $F' \supseteq F/v$ and wX^ω are disjoint.

To prove the converse implication, suppose F to be not nowhere dense in W^ω, that is, according to Lemma 4.3 and Corollary 4.2 there is some $u \in W^*$ such that $\mathcal{C}(F)/uw \supseteq \mathcal{C}(W^\omega)/w \supseteq \mathcal{C}(W^\omega)$, for some $w \in W^*$. Hence, $\mathbf{A}(F) \supseteq vW^*$ and there is no $v \in W^*$ with $F \cap uwvX^\omega = \emptyset$.

2. In view of Lemma 4.3 from $\mathcal{C}(F) \subseteq \mathcal{C}(W^\omega)$ we deduce that $F'' = \bigcup_{u\in X^*} F/u$ is also nowhere dense in W^ω provided F is nowhere dense in W^ω. Now the proof proceeds as in 1. The converse implication of the second part is an immediate consequence of the first part. □

The followng results comes from Staiger [41]:

Corollary 4.5. *A finite-state set $F \in X^\omega$ is nowhere dense if and only if there is a $w \in X^*$ such that $F \subseteq (X^{|w|} \setminus \{w\})^\omega$.*

Proof. Obviously, $(X^{|w|} \setminus \{w\})^\omega$ is nowhere dense. On the other hand, if F is finite-state and nowhere dense then Lemma 4.4 states that $F \subseteq X^\omega \setminus X^* w X^\omega$ for some $w \in X^*$. Now our assertion follows from $X^\omega \setminus X^* w X^\omega \subseteq (X^{|w|} \setminus \{w\})^\omega$. □

Finally, we introduce the set R_0 as the union of all finite-state nowhere dense sets:

$$R_0 = \bigcup_{w\in X^*} (X^\omega \setminus X^* w X^\omega) = \bigcup_{w\in X^*} (X^{|w|} \setminus \{w\})^\omega.$$

The set R_0 is, therefore, the complement of the set of all disjunctive ω-words over X, that is $R_0 = X^\omega \setminus D$. From Corollary 4.5 it is clear that R_0 is a countable union of nowhere dense sets, hence meager or of the first Baire category. Consequently, D is a residual set. The set R_0, and hence its complement D, prove to be useful in the considerations of product measures.

4.2 Product measure and category

In this section we present a connection between measure and density for regular ω-languages (cf. also Staiger [41]) the proof of which makes use of the complement of the set of all disjunctive ω-words, $X^\omega \setminus D = R_0$. To this end we introduce non-degenerated product measures $\bar{\mu}$ on X^ω as follows.

We start with a probability measure $\mu : X \to (0,1]$ on the alphabet X and extend it via $\mu(wv) = \mu(w)\mu(v)$ to a Bernoulli measure on X^*. The product measure $\bar{\mu}$ derived from μ is the σ-additive measure on X^ω which satisfies $\bar{\mu}(wX^\omega) = \mu(w)$.

Due to the requirement $\mu(x) > 0$, for all $x \in X$, every nonempty open subset of (X^ω, ϱ) has non-null measure. We obtain a first property.

Property 4.6. *If $\bar{\mu}$ is a non-degenerated product measure on X^ω, then $\bar{\mu}(R_0) = 0$.*

Proof. Obviously, $\bar{\mu}((X^{|w|} \setminus \{w\})^\omega) = 0$, for every $w \in X^*$. □

Corollary 4.7. *For every non-degenerated product measure on X^ω, $\bar{\mu}(D) = 1$.*

Corollary 4.8. *For every finite-state nowhere dense subset $F \subseteq X^\omega$ we have $\bar{\mu}(\mathcal{C}(F)) = 0$.*

We proceed with the topologically more complicated ω-languages in the classes $\mathbf{F}_\sigma$ and $\mathbf{G}_\delta$. To this end let $\mathcal{I}(F) = X^\omega \setminus \mathcal{C}(X^\omega \setminus F)$ be the *interior* of the ω-language F, that is, the largest open subset contained in F.

Lemma 4.9. *Let $F = \bigcup_{i \in \mathbb{N}} F_i$ where every F_i is finite-state and closed. Then $F \supseteq \mathcal{I}(F) \supseteq F \setminus R_0$ and $\bar{\mu}(F) = \bar{\mu}(\mathcal{I}(F))$.*

Proof. For every finite-state closed (and thus regular) set $F_i \subseteq X^\omega$ the set $F_i \setminus \mathcal{I}(F_i)$ is nowhere dense and regular, and thus a subset of R_0. □

By complementation we obtain

Lemma 4.10. *Let $E = \bigcap_{i \in \mathbb{N}} E_i$ where every E_i is finite-state and open. Then $E \cup R_0 \supseteq \mathcal{C}(E) \supseteq E$ and $\bar{\mu}(E) = \bar{\mu}(\mathcal{C}(E))$.*

Since regular $\mathbf{F}_\sigma$-sets satisfy the conditions of Lemma 4.9 (cf. Staiger and Wagner [48]) and, therefore, regular $\mathbf{G}_\delta$-sets satisfy the conditions of Lemma 4.10, we obtain:

Corollary 4.11. *A regular G_δ-set E is nowhere dense if and only if it is a null set.*

Proof. If E is nowhere dense and regular then $\bar{\mu}(\mathcal{C}(E)) = 0$ according to Corollary 4.8. If E is a null set then from Lemma 4.10 we have $\bar{\mu}(\mathcal{C}(E)) \leq \bar{\mu}(E) + \bar{\mu}(R_0) = 0$. Hence, $\mathcal{C}(E)$ is a closed null set and, thus, nowhere dense. □

This corollary is not valid under the assumptions of Lemma 4.10 because any closed set F is the intersection of regular open sets, but need not be a null set if it is nowhere dense. Neither this corollary is valid for arbitrary regular sets as the example X^*x^ω $(x \in X)$ of a dense null set shows.

Now, we can present the measure-category result for arbitrary regular ω-languages.

Theorem 4.12. *A regular set $F \subseteq X^\omega$ is of first Baire category if and only if it is a null set for any non-degenerated product measure $\bar{\mu}$.*

Proof. We show that $F \subseteq R_0$ if and only if $\bar{\mu}(F) = 0$. Since F is regular, it follows (cf. Thomas [50]) that there are regular $\mathbf{F}_\sigma$-sets F_i and regular $\mathbf{G}_\delta$-sets E_i such that

$$F = \bigcup_{i=1}^{n} (F_i \cap E_i).$$

From this one has

$$F \subseteq \bigcup_{i=1}^{n} (F_i \setminus \mathcal{I}(F_i)) \cup \bigcup_{i=1}^{n} (\mathcal{I}(F_i) \cap E_i).$$

Then the first union is a subset of R_0, and the second is a regular $\mathbf{G}_\delta$-set. Hence, by Corollary 4.11, the latter is a null set if and only if it is contained in R_0, which proves our assertion. □

Thus we have seen that for an arbitrary finite alphabet X containing at least two letters the ω-language of disjunctive sequences is large as well in the sense of Baire category in Cantor space as in the sense of product measure on X^ω. Moreover, every non-disjunctive sequence is contained in a certain regular ω-language which is nowhere dense in X^ω or, equivalently, has null measure. We shall return to this item later when we will measure the density of regular ω-languages containing a certain non-disjunctive sequence $\mathbf{x}$.

5 How Large is the Set of Disjunctive Real Numbers?

We shall study the size of disjunctive sequences by means of their associated real numbers. The first natural question is the following:[7] Is disjunctiveness a property of real numbers or of their representations? In other words, if a real number is disjunctive to a certain base, is it disjunctive in any other base? Let us first notice that normality is not base invariant (as it is proved, for example, in Kuipers and Niederreiter [24]),[8] but, as it was proved in Calude and Jürgensen [7], randomness is invariant under the change of base (see also Calude [6]).

Two real numbers $a, b > 1$ are *equivalent* if $a^n = b^m$, for some positive integers n, m. The following result is stated in Hertling [19]; its proof is based on results in El-Zanati and Transue [16] and Schmidt [40].

Theorem 5.1. *1. Every real number which is disjunctive to a base b is disjunctive to every base a equivalent to b. 2. Assume a and b are not equivalent bases. Then the set of real numbers which are disjunctive to base a but not disjunctive to base b has the cardinality of the continuum.*

Theorem 5.1 gives us no indication about any "concrete examples" of reals that are disjunctive in one base and not disjunctive in another base.[9] The following result, due to Hertling [19], remedies this situation.

[7] This question was raised by H. Jürgensen and solved by Hertling [19].

[8] A real x is simply normal to base b if every digit $0 \leq d < b$ appears in the sequence $s_b(x)$ with the asymptotic frequency b^{-1}. Let $a, b \geq 2$ and $n \geq 1$ be integers and assume that $a \neq b^n$, for all n. The set of real numbers which are simply normal to base a but not simply normal to base b has the cardinality of the continuum; see Hertling [18, 19]

[9] It is not known whether numbers like e, π, $\ln 2$ are or not disjunctive.

Theorem 5.2. *Let $b \geq 2$ be an integer. The number*

$$\sum_{i=0}^{\infty} b^{-i!-i}$$

is not disjunctive to base b, but is disjunctive to all bases a which are not equivalent to b and are divisible by all prime divisors of b.

From now on let us denote by $\mathcal{D}$ the set of all numbers which are disjunctive in any base.[10] Is $\mathcal{D}$ empty?

Let x be a real number in [0, 1] written in base b, and $x_{(n)} \in \mathbf{Q}$[11] have precisely n digits, namely the first n digits of x (completed with zeros if necessary).

If the word u appears (without overlaps) exactly k times in $x_{(n)}$, put

$$p_{u,n}(x) = \frac{k|u|}{n}.$$

Let

$$p_u^-(x) = \liminf_{n\to\infty} p_{u,n}(x), \quad p_u^+(x) = \limsup_{n\to\infty} p_{u,n}(x).$$

Usually, we say that u appears with probability p in x if

$$p_u^-(x) = p_u^+(x) = p.$$

According to the law of large numbers (see Oxtoby [32]), almost every real number from $[0, 1]$ is normal, i.e. every word appears with its "natural" probability. So, for example, the ones appear with probability 1/2, the word 0010 appears with probability $1/2^4$, if base 2 is used. This property is valid for almost all numbers, but not exactly all of them.[12] In the sense of Baire categories, how do most numbers behave?[13] The answer comes from Calude and Zamfirescu [8]:

[10] These numbers are called *absolutely disjunctive.*

[11] The set of rationals is denoted by $\mathbf{Q}$.

[12] To get a quick idea of this phenomenon let us prove, with Borel [1], the following weaker result (compare with Property 4.6): *Almost all real numbers, when expressed in any base, contain every possible digit or possible word of digits.* Indeed, let $b \geq 2$. Notice that for all $a \in X_b$ and $u \in X_b^*$, u does not contain the digit a if and only if $u \in (X_b \setminus \{a\})^*$. Accordingly, for every $k > 0$

$$N(k) = \#\{u \in X_b^* \mid |u| \leq k \wedge u \text{ doesn't contain } a\} = \frac{(b-1)^{k+1}-1}{b-2},$$

and

$$\frac{N(k)}{\#\{u \in X_b^* \mid |u| \leq k\}} = \frac{((b-1)^{k+1}-1)(b-1)}{(b^{k+1}-1)(b-2)},$$

so

$$\lim_{k\to\infty} \frac{N(k)}{\#\{u \in X_b^* \mid |u| \leq k\}} = 0.$$

We may now write the formula:

$$\lim_{k\to\infty} \frac{\#\{u \in X_b^* \mid |u| \leq k, u \text{ does contain } a\}}{\#\{u \in X_b^* \mid |u| \leq k\}} = 1$$

which shows that almost all reals, when expressed in any scale $b \geq 2$, contain every possible digit $a \in X_b$. The case of words of digits can be easily settled just by working with a large enough base. For instance, if the word 957 never occurs in the ordinary decimal for some number, then the *digit* 957 never occurs in base 1000.

[13] Here "most" means "those in a residual set", i.e. "all, except those in a set of first category".

Theorem 5.3. *For most numbers $x \in [0,1]$, using any base and choosing any word u written in the same base,*

$$p_u^-(x) = 0 \quad \textit{and} \quad p_u^+(x) = 1.$$

Proof. Choose arbitrarily the base b and the word u written in base b.

Let, for some $\alpha \in (0,1) \cap \mathbf{Q}$,

$$\mathcal{R}^+_{\alpha,n} = \{x \in [0,1] \mid \exists m(m \geq n \wedge p_{u,m}(x) \geq \alpha)\}.$$

We claim that $\mathcal{R}^+_{\alpha,n}$ contains an open set dense in $[0,1]$. Indeed, choose $y \in [0,1)$ and $\varepsilon > 0$ arbitrarily. Let $q \geq n$ satisfy $b^{-q} < \varepsilon$. To the digits of $y_{(q)}$ we add the word u as many times as needed in order to get a number $z = z_{(m)}$ (with $z_{(n)} = y_{(n)}$) satisfying $p_{u,m}(z) \geq \alpha$. Then the whole interval $(z, z + b^{-m})$ lies in $\mathcal{R}^+_{\alpha,n}$ and each of its points has distance at most $b^{-m} < b^{-q} < \varepsilon$ from y. The claim is proven.

Analogously,

$$\mathcal{R}^-_{\alpha,n} = \{x \in [0,1] \mid \exists m(m \geq n \wedge p_{a,m}(x) \leq \alpha)\}$$

contains an open dense set. Of course, $\mathcal{R}^-_{\alpha,n}$ and $\mathcal{R}^+_{\alpha,n}$ depend on b and u. Thus,

$$\mathcal{R}_{u,b} = \bigcap_{\alpha,n} (\mathcal{R}^-_{\alpha,n} \cap \mathcal{R}^+_{\alpha,n})$$

is residual in $[0,1]$.

Notice that

$$\bigcap_n \mathcal{R}^-_{\alpha,n} = \{x \in [0,1] \mid p_u^-(x) \leq \alpha\}, \ \bigcap_{\alpha,n} \mathcal{R}^-_{\alpha,n} = \{x \in [0,1] \mid p_u^-(x) = 0\},$$

and similarly for $+$ instead of $-$. Therefore $\cap_{u,b}\mathcal{R}_{u,b}$ is exactly the set of real numbers the theorem speaks about, and it is residual in $[0,1]$. □

A consequence of Theorem 5.3 is that most numbers in $[0,1]$ are absolutely disjunctive. As another immediate consequence we get a result due to Oxtoby and Ulam ([33], p. 877):

Corollary 5.4. *The law of large numbers is false in the sense of category.*

Proof. Indeed, the set of all numbers $x \in [0,1]$ such that in their dyadic development the digits 0 and 1 appear with probability one-half lies in the complement of the residual set from Theorem 5.3. □

Random numbers are transcendental, as they are non-computable Calude [6]. This argument does not work for disjunctive numbers as some of these numbers are computable. Using a result of Staiger [42][14] Jürgensen and Thierrin [21] have proved:

[14] Stating that the set $\{\mathbf{x}\}$ is an ω-language accepted by a deterministic finite acceptor provided $v_b(\mathbf{x})$ is rational.

Theorem 5.5. *If $\mathbf{x} \in X_b^\omega$ is disjunctive, then $v_b(\mathbf{x})$ is not rational.*

There are uncountably many disjunctive sequences $\mathbf{x}$ having $v_b(\mathbf{x})$ transcendental as the set of random sequences has measure 1, Calude [6]. The proof of the following theorem, due to Jürgensen and Thierrin [21], explicitly constructs an uncountable class of transcendental disjunctive numbers. A real number x is a *Liouville number* if it is not rational and if for each positive integer n there exist two integers (depending upon n) $q_n > 1$ and p_n such that $|x - \frac{p_n}{q_n}| < q_n^{-n}$ holds true. Liouville numbers are typical examples of transcendental numbers Hardy and Wright [17].

Theorem 5.6 (Jürgensen and Thierrin [21]). *There are uncountably many disjunctive Liouville numbers.*

Proof. Consider a sequence $\{w_i\}_{i\geq 1}$ of words $w_i \in X_b^* \setminus \{\epsilon\}$ containing exactly one occurrence of each non-empty word. Let $t_i = |w_i|$ and $r_i = v_b(w_i)$, that is

$$r_i = \sum_{j=1}^{t_i} x_{i_j} b^{-j},$$

where $w_i = x_{i_1} x_{i_2} \cdots x_{i_{t_i}}$.

Define

$$\lambda_1 = 0,\ \lambda_i = \sum_{j=1}^{i-1} t_j, \quad \text{for } j > 1,$$

and

$$x = \sum_{i=1}^{\infty} r_i b^{-\lambda_i!},$$

and let $\mathbf{y} = s_b(x)$ be the base-b expansion of x. Clearly, $\mathbf{y}$ is disjunctive. We will show that x is transcendental by proving that x is a Liouville number. Given the positive integer n construct $\lambda_k > n$ and minimum with this property, and define $q_n,\ p_n$ by

$$\frac{p_n}{q_n} = \sum_{i=1}^{k} r_i b^{-\lambda_i!},\ q_n = b^{\lambda_k!+t_k}.$$

The difference $x - \frac{p_n}{q_n}$ can be evaluated as follows:

$$\begin{aligned}
0 < x - \frac{p_n}{q_n} & \\
&= x - \sum_{i=1}^{k} r_i b^{-\lambda_i!} \\
&= \sum_{i=k+1}^{\infty} r_i b^{-\lambda_i!} \\
&< \frac{b-1}{b^{\lambda_{k+1}!+1}} \sum_{j=1}^{\infty} b^{-j} \\
&= b^{-\lambda_{k+1}!} \\
&= \frac{1}{b^{(\lambda_{k+1}-1)!\lambda_{k+1}}}.
\end{aligned}$$

One has $\lambda_k < \lambda_{k+1}$; if we assume, in addition, that

$$\lambda_k! + t_k < (\lambda_{k+1} - 1)!, \tag{11}$$

then

$$x - \frac{p_n}{q_n} < \frac{1}{q_n^{\lambda_k}} < q_n^{-N},$$

for all $N < \lambda_k$. It remains to prove that the inequality (11) is satisfied for almost all positive integers k. Indeed, observe that the sequence $\{\lambda_n\}_{n\geq 1}$ is strictly increasing, so $\lambda_k \geq 2$ for almost all k. From the construction of the sequence $\{w_i\}_{i\geq 1}$ it follows that $t_k \geq 2$ for almost all k. Therefore,

$$t_k < t_k \lambda_k! \leq (\lambda_k + t_k - 2)\lambda_k!,$$

so

$$\lambda_k! + t_k < \lambda_k!(\lambda_k + t_k - 1) \leq (\lambda_k + t_k - 1)! = (\lambda_{k+1} - 1)!,$$

for almost all k. □

Corollary 5.7. *There exist uncountable many disjunctive sequences* $\mathbf{x} \in X_b^\omega$ *such that* $v_b(\mathbf{x})$ *is transcendental.*

A stronger results holds true:

Theorem 5.8. *Most Liouville numbers are absolutely disjunctive.*

Proof. Since the residual set from Theorem 5.3 is a subset of $\mathcal{D}$, most numbers from $[0,1]$ are in $\mathcal{D}$. But most reals are Liouville numbers, cf. Oxtoby [32], Priestley [38]. □

So, most numbers from $[0,1]$ lie in $\mathcal{D}$. Simultaneously, $\mathcal{D}$ contains all elements of $[0,1]$ obeying the law of large numbers and has therefore measure 1. This suggests that $\mathcal{D}$ may contain nearly all elements of $[0,1]$. But what means "nearly all"? To answer this question we shall make use, following Calude and Zamfirescu [8], of the notion porosity studied by Dolzhenko [15]. A set $M \subset [0,1]$ is said to be *porous at* $x \in [0,1]$ if there is a number $\beta > 0$ and a sequence of points $\{x_n\}_{n=1}^{\infty}$ converging to x such that for large enough n,

$$(x_n - \beta|x - x_n|, \quad x_n + \beta|x - x_n|) \cap M = \emptyset.$$

Further, M is called *porous* if it is porous at each of its points, and it is called σ*-porous* if it is a countable union of porous sets. We say that *nearly all* points of $[0,1]$ enjoy property P if the set of points not enjoying P is σ-porous Zamfirescu [53]. By Lebesgue's Density Theorem,[15] every porous set has measure zero and therefore a set containing nearly all elements is large from both the measure and the category points of view.[16]

[15] See Oxtoby [32].

[16] The complement of a null set of first category may well not contain nearly all elements, as Zajíček [52] proved.

Theorem 5.9. *Nearly all numbers in* $[0,1]$ *are absolutely disjunctive.*

Proof. Let u be a word written in base b. Let $\mathcal{P}_{u,b}$ be the set of all numbers in $[0,1]$ which, written in base b, do not contain u. We show that $\mathcal{P}_{u,b} \setminus \{1\}$ is porous.

Let $y \in \mathcal{P}_{u,b} \setminus \{1\}$. Consider the arbitrary natural number n. Using the notation from Theorem 5.3, $y - y_{(n)} \leq b^{-n}$. Add u to the digits of $y_{(n)}$ and get another number $z = z_{(m)}$ with $m = n + |u|$. Clearly,

$$(z, z + b^{-m}) \cap \mathcal{P}_{u,b} = \emptyset.$$

For the midpoint z_m of $(z, z + b^{-m})$, we have $|z_m - y| < b^{-n}$ and

$$b^{-m}/2 = 2^{-1}b^{-|u|}b^{-n} > 2^{-1}b^{-|u|}|z_m - y|,$$

whence

$$(z_m - 2^{-1}b^{-|u|}|y - z_m|,\ \ z_m + 2^{-1}b^{-|u|}|y - z_m|) \cap \mathcal{P}_{u,b} = \emptyset.$$

while $z_m \to y$ when $n \to \infty$. Hence $\mathcal{P}_{u,b} \setminus \{1\}$ is porous and $\cup_{u,b}\mathcal{P}_{u,b}$, the set in the statement, is σ-porous. □

The above results hold true even constructively. To each word $w \in X_b^*$ we associate the open interval $I_{b,w} = (v_b(w), v_b(w) + b^{-|w|}) \subset [0,1]$ where $v_b(w_1 w_2 \cdots w_n) = \sum_{i=1}^{n} w_i b^{-i}$ in case $w = w_1 w_2 \ldots w_n$. The family $\{I_{b,w}\}_{w \in X_b^*}$ is a base for the natural topology on $[0,1]$. First we get a constructive version of a residual set (i.e. of a set that contains the intersection of a countable family of open dense sets). To this aim we require that the family of open dense sets is enumerated by a recursively enumerable (r.e.) set, and each basic open set $I_{b,x}$ intersects the family in an effectively computable point. We are led to the following definition: A set R is *constructively residual* if there exists an r.e. set $E \subset \mathbb{N} \times \mathbb{N}^* \times \mathbb{N}$, and a recursive function $f : \mathbb{N}^* \times \mathbb{N} \to \mathbb{N}^*$ such that the following two conditions hold true:

1. The set
$$\bigcap_{m=1}^{\infty} \left(\bigcup_{(b,x,m) \in E} I_{b,x} \right)$$
is contained in R.
2. For all $b \geq 2, m \geq 1,$ and $x \in B_b^+$ we have $x \subset f(x,m),\ (b, f(x,m), m) \in E.$

A constructively residual set is residual, but the converse is false. The statement "constructively, the typical number has, or most numbers have, property P" means that the set of all numbers with property P is constructively residual. In Calude and Zamfirescu [9] one obtains the following:

Theorem 5.10. *Constructively, for most numbers* $x \in [0,1]$, *using any base* b *and choosing any word* $v \in B_b^+$

$$p_{b,v}^{-}(x) = 0 \text{ and } p_{b,v}^{+}(x) = 1.$$

Corollary 5.11. *Constructively, a typical number does not obey the law of large numbers.*

Corollary 5.12. *Constructively, the typical Liouville number is absolutely disjunctive.*

To get a constructive version of Theorem 5.9 we need first a constructive version of Lebesgue Density Theorem. Call a set $M \subset [0,1]$ *constructively megaporous* if there exist a base $b \geq 2$, a rational number $r \in (0,1)$ and a recursive function $f : X_b^* \to X_b^*$ such that each interval $I_{b,u}$ of length less than r contains a subinterval $I_{b,f(u)}$ disjoint from M and having length greater than $rb^{-|u|}$. An r.e. union of constructively megaporous sets is called *constructively σ-megaporous.* More precisely, M is constructively σ-megaporous if $M = \bigcup_{n=1}^{\infty} M_n$, and there exist two recursive functions $T : \mathbb{N} \times \mathbb{N}^* \to \mathbb{N}^*$, $R : \mathbb{N} \to \mathbf{Q}$ such that M_n is constructively megaporous under $T(n,.)$ and $R(n)$. We say that "constructively, nearly every point of $[0,1]$ enjoys property P" if the set of points not enjoying P is constructively σ-megaporous.

Following Martin-Löf [27] (see Calude [6]), a set $S \subset [0,1]$ is constructively null (with respect to Lebesgue measure μ) if there exists a base $b \geq 2$ and an r.e. set $G \subset X_b^+ \times \mathbb{N}$ such that

$$S \subset \bigcap_{n=1}^{\infty} \left(\bigcup_{(x,n) \in G} I_{b,x} \right),$$

and

$$\lim_{n \to \infty} \mu \left(\bigcup_{(x,n) \in G} I_{b,x} \right) = 0, \text{ constructively.}$$

The following weaker version of Lebesgue Density Theorem, due to Calude and Zamfirescu [9], is useful for our aims.

Theorem 5.13. *Every constructively σ-megaporous set is constructively null.*

Proof. Due to Martin-Löf's Theorem (Calude [6]), there exists a maximal constructively null set. Consequently, it is enough to prove the theorem for constructively megaporous sets.

Let M be constructively megaporous with respect to the base b, the rational r and the recursive function f. To estimate the size of M we will generate, in a recursive way, smaller and smaller coverings of M. We start with an integer n such that $b^{-n} < r$. For a word $w \in B_b^+$ put $E(w) = \{y \in B_b^+ \mid w \subset y, |y| = |f(w)|, \text{ and } y \neq f(w)\}$. The first covering is

$$M \subset \bigcup_{|u|=n} I_{b,u}.$$

The second iteration is

$$M \subset \bigcup_{|u|=n} \bigcup_{v_1 \in E(u)} I_{b,v_1} = \bigcup_{|u|=n} I_{b,u} \setminus I_{b,f(u)}.$$

The measure of this covering is

$$\mu\left(\bigcup_{|u|=n} I_{b,u}\setminus I_{b,f(u)}\right) = \sum_{|u|=n}\mu(I_{b,u}\setminus I_{b,f(u)})$$
$$= \sum_{|u|=n}(b^{-|u|}-b^{-|f(u)|})$$
$$\leq \sum_{|u|=n} b^{-|u|}(1-r)$$
$$= 1-r.$$

In general, a proof by induction shows that

$$M \subset \bigcup_{|u|=n}\bigcup_{v_1\in E(u)}\cdots\bigcup_{v_k\in E(v_{k-1})}\bigcup_{v_{k+1}\in E(v_k)} I_{b,v_{k+1}}$$
$$= \bigcup_{|u|=n}\bigcup_{v_1\in E(u)}\cdots\bigcup_{v_k\in E(v_{k-1})} I_{b,v_k}\setminus I_{b,f(v_k)}$$

and

$$\mu\left(\bigcup_{|u|=n}\bigcup_{v_1\in E(u)}\cdots\bigcup_{v_k\in E(v_{k-1})}\bigcup_{v_{k+1}\in E(v_k)} I_{b,v_{k+1}}\right) < (1-r)^{k+1}.$$

We conclude that M is constructively null with respect to the r.e. family $G = \{(w,n) \in B_b^+ \times \mathbb{N} \mid w \in F_n, n = 1,2,\ldots\}$, where $F_0 = \{u \in B_b^+ \mid |u| = n\}$ and $F_{k+1} = \{u \in B_b^+ \mid u \in E(w),\text{ for some } w \in F_k\}$. □

We conclude this section with one more result in Calude and Zamfirescu [9]:

Theorem 5.14. *Constructively, nearly every number is absolutely disjunctive.*

Proof. Let $\gamma : \{(b,w) \mid b \geq 2 \wedge w \in B_b^+\} \to \mathbb{N}$ be a recursive bijection, and define the recursive functions $T(n,u) = uw$, $R(n) = b^{-|w|} - 1$, whenever $n = \gamma(b,w)$. Again, if $n = \gamma(b,w)$, we put $L_n = \{0 \leq x \leq 1 \mid w \text{ is not contained in } s_b(x)\}$. It is seen that $[0,1]\setminus\mathcal{D} = \bigcup_{i=1}^{\infty} L_i$, and each L_n is constructively megaporous with respect to the base b, the recursive function $T(n,.)$ and the rational $R(n)$. □

Theorem 5.14 is stronger than Theorem 5.9 as, for instance, constructively null sets are much smaller than classical null sets: the union of all null sets coincides with the whole space, while, the union of all constructively null sets is a constructively null set.

6 How Complex Are Disjunctive Sequences?

6.1 A computational complexity approach

At the top, disjunctive sequences $\mathbf{x}$ can be random, non-random but non-recursive, recursive, but arbitrarily complex; they form a class of measure one Calude [6].

To construct, following Calude and Yu [10], arbitrarily complex recursive disjunctive sequences we rely on Rabin's Theorem (see, for instance, Theorem 3.5 in Calude [4]).[17]

Theorem 6.1. *There exist recursive, arbitrarily complex, disjunctive sequences.*

Proof. Consider a primitive recursive enumeration ε of all non-empty words over X and a recursive function B mapping positive integers into positive integers. Assume that B grows as fast as Ackermann's function Calude [4]. Fix two letters, say σ_1, σ_2 in X and let f be a recursive function mapping positive integers into $\{\sigma_1, \sigma_2\}$ such that for every $\varphi_i = f$, $\Phi_i(n) > 2B(n)$, for almost all n. Construct the sequence

$$\mathbf{x} = f(1)\varepsilon(1)f(2)\varepsilon(2)\cdots f(n)\varepsilon(n)\cdots.$$

Clearly, $\mathbf{T}(\mathbf{x}) = X^*$. Let $\mathbf{x}(n)$ be the prefix of length n of $\mathbf{x}$. Then for every integer $n \geq 2$,

$$f(n) = \psi(\mathbf{x}(\sum_{i=1}^{n-1} \mid \varepsilon(i) \mid +n)),$$

where $\psi(w)$ returns the last letter of the word w. Obviously, if $\varphi_j(n) = x_n$, (x_n is the nth term of $\mathbf{x}$), then $\Phi_j(n) > B(n)$, for almost all n. □

6.2 A language-theoretic approach

At the bottom, the complexity of a sequence $\mathbf{x}$ will be measured by the complexity of the language $\mathbf{A}(\mathbf{x})$ consisting of all prefixes of $\mathbf{x}$; these languages can be context-sensitive, but not context-free.

One can prove that the language of all prefixes of the sequence consisting of all words over the binary alphabet arranged in quasi-lexicographical order is context-sensitive (see Calude and Yu [10]).[18] The fact that this complexity is the best possible will be proven below.

Lemma 6.2. *A sequence $\mathbf{x}$ is ultimately periodic if and only if its set of prefixes $\mathbf{A}(\mathbf{x})$ is context-free.*

Proof. If $\mathbf{x} = wv^\omega$ for some words $w, v \in X^*$ then $\mathbf{A}(\mathbf{x}) = \mathbf{A}(wv^*)$ is even regular.

Conversely, since $\mathbf{A}(\mathbf{x})$ is infinite, according to the pumping lemma for context-free languages it contains an infinite family $(w'w^n vu^n u')_{n\in\mathbb{N}}$. Thus $\mathbf{x} = w'w^\omega$ if w is not the empty word or otherwise $\mathbf{x} = w'vu^\omega$. □

[17] *For every Blum space (φ_i, Φ_i) and for every recursive function B, a two-valued recursive function f can be effectively constructed such that, for every j with $\varphi_j = f$, one has $\Phi_j(n) > B(n)$, for almost all n.*

[18] In fact, there exist infinitely many disjunctive sequences $\mathbf{x} \in X^\omega$ such that $\mathbf{A}(\mathbf{x})$ is context-sensitive.

As consequences we deduce:

Corollary 6.3. *For every disjunctive sequence* $\mathbf{x} \in X^\omega$, $\mathbf{A}(\mathbf{x})$ *is not context-free.*

Corollary 6.4. *If* $\mathbf{x} \in X^\omega$ *is disjunctive, then* $\mathbf{A}(\mathbf{x})$ *contains no infinite context-free language.*

6.3 Subword complexity

We introduce a concept of complexity of infinite sequences $\mathbf{x}$ which is intimately related to disjunctive ω-words. This concept is based solely on the sets of subwords $\mathbf{T}(\mathbf{x})$. It turns out that the subword complexity $\tau(\mathbf{x})$ of a word $\mathbf{x} \in X^\omega$ is also closely related to the entropy of the regular ω-languages containing $\mathbf{x}$.

For a language $W \subseteq X^*$ let

$$\mathbf{s}_W(n) = \operatorname{card} W \cap X^n$$

be its *structure function* (cf. Kuich [23]), and

$$H_W = \limsup_{n \to \infty} \frac{\log_{cardX} \mathbf{s}_W(n)}{n}$$

be its *entropy*. Define $\mathbf{s}_F = \mathbf{s}_{\mathbf{A}(F)}$ and $H_F = H_{\mathbf{A}(F)}$, for $F \subseteq X^\omega$.

We continue with more definitions. Let

$$\mathbf{T}_\infty(\mathbf{x}) = \{w \mid w \in X^* \wedge \forall n \exists v \exists \beta (|v| \geq n \wedge vw\beta = \mathbf{x})\} \tag{12}$$

be the set of subwords occurring infinitely often in $\mathbf{x}$.

We call $\tau(\mathbf{x}) = H_{\mathbf{T}(\mathbf{x})}$ the *subword complexity* of the word $\mathbf{x} \in X^\omega$. Then the following identity holds (cf. Staiger [46]):

$$H_{\mathbf{T}(\mathbf{x})} = H_{\mathbf{T}_\infty(\mathbf{x})} = \inf\{H_F \mid F \text{ is finite-state } \wedge \mathbf{x} \in F\} \tag{13}$$

An ω-word is disjunctive if and only if $\mathbf{T}(\mathbf{x}) = \mathbf{T}_\infty(\mathbf{x}) = X^*$.

Proposition 6.5. *Let* F *be a finite-state* ω*-language. Then* $\tau(\mathbf{x}) \leq H_F$*, for every* $\mathbf{x} \in F$*.*

The following result comes from Staiger [46].

Proposition 6.6. *If* $F \neq \emptyset$ *is a closed finite-state* ω*-language then there is a recursive (as a function mapping* $\mathbb{N}$ *to* X*) sequence* $\mathbf{x} \in F$ *satisfying* $\tau(\mathbf{x}) = H_F$*.*

Now, in a way similar to Lemma 4.4 we obtain a relationship between density, entropy and subwords:

Theorem 6.7. *Let* $E \subseteq X^\omega$ *be such that* $\mathcal{C}(E) = \mathcal{C}(W^\omega)$*, for some nonempty regular language* $W \subseteq X^*$*, and let* F *be a finite-state subset of* $\mathcal{C}(E)$*.*

Then the following conditions are equivalent:

1. *F is nowhere dense in E.*
2. *$H_F < H_E$.*
3. $\lim\limits_{n\to\infty} \frac{\mathbf{s}_F(n)}{\mathbf{s}_E(n)} = 0$
4. *$\exists w(w \in W^* \wedge F \subseteq \mathcal{C}(E) \setminus X^* w X^\omega)$.*
5. *$\forall \mathbf{x}(\mathbf{x} \in F \Rightarrow \tau(\mathbf{x}) < H_E)$.*

Proof. Theorem 21 in Merzenich and Staiger [29] shows that the first three conditions are pairwise equivalent. Moreover, Lemma 4.4 proves the equivalence of *1.* and *4.*, and the above Propositions 6.5 and 6.6 show the equivalence of *2.* and *5.*. □

7 Disjunctive Runs

7.1 Elementary properties

In this section we study a canonical generalization of disjunctive sequences, namely disjunctive runs in finite automata. We need some further notation from ω-languages. Let M, N denote languages over X, i.e. $M, N \subseteq X^*$.

An *automaton* A (over X) is a tuple $A = (Q_A, E_A, i_A)$ of a finite set Q_A of *states*, $Q_A \cap X = \emptyset$, an *initial state* $i_A \in Q_A$, and a relation $E_A \subseteq Q_A \times X \times Q_A$ of *labeled arcs*. *A run* r in A *from* q *to* q' is a word $r = e_1 \ldots e_n \in {E_A}^*$ with $e_i = (q_i, a_i, q'_i) \in E_A$ such that $q'_i = q_{i+1}$ holds for $1 \le i < n$ with $q_1 = q$ and $q'_n = q'$. *Run(A)* is the set of all finite runs in A. An *infinite run* $\mathbf{r}$ in A is an element of adh $Run(A) \subseteq {E_A}^\omega$. $Run^\omega(A)$ is the set of all infinite runs in A starting from i_A. As (infinite) runs are special words in $E^\infty, \mathbf{r}[k], \mathbf{r}[k,l], r[k]$, and $r[k,l]$ are already defined.

The *labeling* λ: $E_A \to X$ is the projection of a labeled arc onto its second coordinate, $\lambda((q,a,q')) = a$. λ is canonically extended to a homomorphism $\lambda : {E_A}^\infty \to X^\infty$. $Run_{q,q'}(A)$ denotes the set of all finite runs in A from q to q', $L_{q,q'}(A) := \lambda(Run_{q,q'}(A))$. Note that $\epsilon \in L_{q,q}(A)$, as the empty run runs from q to q for any q. $M \subseteq X^*$ is called *recognizable* if there is an automaton A and a set Q^f of *final states*, $Q^f \subseteq Q_A$, such that $M = \bigcup_{q' \in Q^f} L_{i_A,q'}(A)$; *Rec* denotes the class of all recognizable languages. Analogously, *Reg* denotes the set of all regular languages. *Rec* = *Reg* is known as Kleene's Main Theorem in automata theory, cf. Kleene [22].

An automaton A is called *reachable* if for any state $q \in Q_A$ there is a run in A from i_A to q. One may usually drop all non-reachable states. In the sequel, any automaton will be reachable. An automaton A is called *deterministic* if $(q,a,q_1) \in E_A$ and $(q,a,q_2) \in E_A$ implies $q_1 = q_2$, for all $q, q_1, q_2 \in Q_A$ and all $a \in X$. A (usually non-deterministic) automaton A with an attached set $Q^f \subseteq Q_A$ of final states is called *trim* if for any state $q \in Q_A$ there is a finite run r from i_A to some final state $q' \in Q^f$ that runs through q. A is called *strongly connected* if for any two states $q, q' \in Q_A$ there is a run in A from q to q'. Obviously, any recognizable language can be recognized by some trim automaton by throwing away all states that are not reachable or that do not lead to any final state by any run in A, and for any regular $*$-closed language M (i.e., a language satisfying the equality $M^* = M$) there is a strongly connected automaton A such that $M = L_{i_A,i_A}(A)$.

We say that an infinite run $\mathbf{r} = (e_i)_{i \in \mathbb{N}}$, $e_i = (q_i, a_i, q'_i)$, *touches a state* q *infinitely often* if $q = q_i$ for infinitely many $i \in \mathbb{N}$. $Q^\omega(\mathbf{r})$ denotes the set of states touched infinitely often by $\mathbf{r}$. An ω-language $F \subseteq X^\omega$ is called *recognizable* if there is an automaton A and a set $Q^f \subseteq Q_A$ of final states such that $F = \{\mathbf{x} \in X^\omega \mid \exists\ \mathbf{r} \in Run^\omega(A)\ (\lambda(\mathbf{r}) = \mathbf{x} \wedge Q^\omega(\mathbf{r}) \cap Q^f \neq \emptyset)\}$. Rec^ω denotes the class of recognizable ω–language. Let Reg^ω denote the class of regular ω–languages, see chapter 1. By famous results of Büchi [3], McNaughton [28], and Muller [30], $Rec^\omega = Reg^\omega$.

In what follows we shall use disjunctivity as another acceptance criterion for infinite runs. An infinite run $\mathbf{r}$ is called disjunctive if any finite run that may be used infinitely often by $\mathbf{r}$ has to be used at least once (and thus infinitely often) by $\mathbf{r}$. More formally, $\mathbf{r}$ is *disjunctive* if for all $q \in Q^\omega(\mathbf{r})$ and for all finite runs r in A from q r is an infix of $\mathbf{r}$. We talk about λ-*disjunctivity* if $\mathbf{r}$ may not use all possible sub-runs but $\lambda(\mathbf{r})$ must contain all labelings of all possible sub-runs. Formally, $\mathbf{r}$ is called λ-*disjunctive* if for all $q \in Q^\omega(\mathbf{r})$ and for all finite runs r in A from q $\lambda(r) \in \lambda(\mathbf{T}(\mathbf{r}))$ holds. This leads to (λ-)disjunctive languages of automata:

$$L^d(A) = \{\mathbf{x} \in X^\omega \mid \exists \mathbf{r} \in Run^\omega(A)\ (\mathbf{r} \text{ is disjunctive} \wedge \mathbf{x} = \lambda(\mathbf{r}))\}$$

is the *disjunctive language* of A,

$$L^{\lambda-d}(A) = \{\mathbf{x} \in X^\omega \mid \exists \mathbf{r} \in Run^\omega(A)\ (\mathbf{r} \text{ is } \lambda\text{-disjunctive} \wedge \mathbf{x} = \lambda(\mathbf{r}))\}$$

is the λ-*disjunctive language* of A.[19]

Finally, Rec^d and $Rec^{\lambda-d}$ denote the classes of disjunctive and λ-disjunctive languages.

Here is a list of examples.[20]

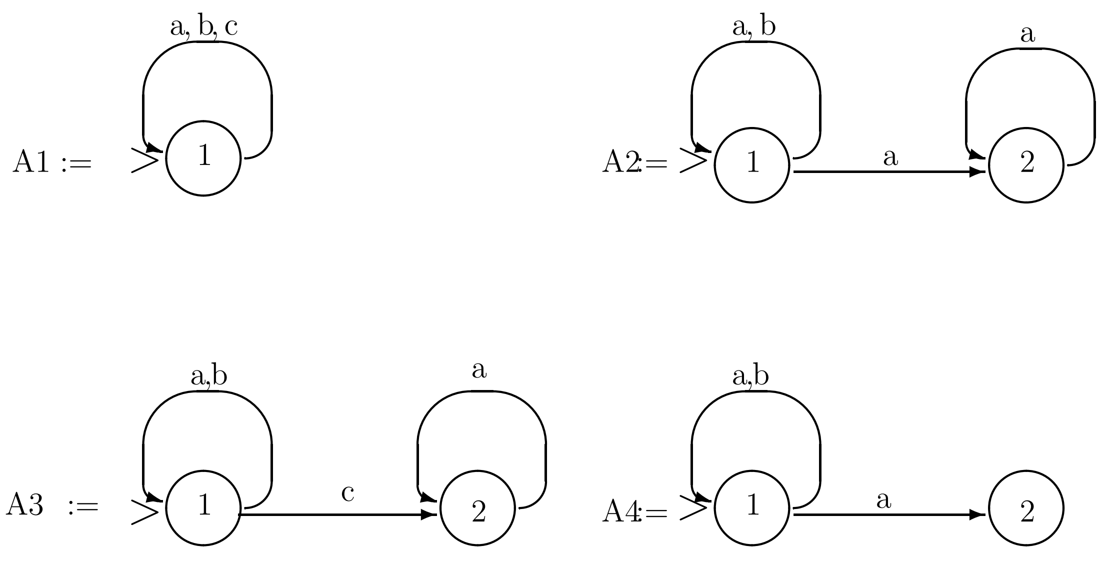

Figure 1. Some automata

[19] In the literature, the term "disjunctive language" of an automaton is also used regarding properties of the syntactic monoid. However, in this paper, a disjunctive language is simply the λ-image of the set of all disjunctive runs of an automaton.

[20] We represent automata, as usual, by graphs; $>$ denotes the initial state.

Examining Figure 1 we immediately get:

- $L^d(A_1) = L^{\lambda-d}(A_1) = D_{\{a,b,c\}}$,
- $L^d(A_2) = \{a,b\}^* a^\omega$, as any disjunctive run must reach state 2 the arc $1 \xrightarrow{a} 2$ must be used,
- $L^{\lambda-d}(A_2) = \{a,b\}^* a^\omega \cup D_{\{a,b\}}$, as a λ-disjunctive run may stay with state 1: any labeling string can be used without leaving state 1,
- $L^d(A_3) = L^{\lambda-d}(A_3) = \{a,b\}^* c a^\omega$,
- $L^d(A_4) = \emptyset$, as any disjunctive run in A_4 must use $1 \xrightarrow{a} 2$ and thus cannot be infinite,
- $L^{\lambda-d}(A_4) = D_{\{a,b\}}$.

As can be seen in the above examples (λ)–disjunctive languages may be ω–regular.[21]
A *sub-automaton* B of A generated by $Q_B \subseteq Q_A$ is given by $B = (Q_B, E_B, i_B)$ with some $i_B \in Q_B$ and $E_B = E_A \cap (Q_B \times X \times Q_B)$. A sub-automaton B of A is called *final* (λ-*final*) if there is a disjunctive (λ-disjunctive) run $\mathbf{r}$ in A from i_A such that $Q_B = Q^\omega(\mathbf{r})$. A sub-automaton B of A is final if $E_B \neq \emptyset$ and B is strongly connected and closed (i.e., no run in A starting from a state in B can reach a state in $Q_A \setminus Q_B$). A sub-automaton B of A is λ-final if $E_B \neq \emptyset$ and B is strongly connected and λ-closed (i.e., for any finite run r in A starting from some state q in B there exists a run r' in B with $\lambda(r') = \lambda(r)$). Thus, any final sub-automaton is also λ-final but not necessarily vice versa. $Fin(A)$ and $Fin^\lambda(A)$ denote the (finite) sets of all final and λ-final sub-automata of A. The following (obvious) normal-form-theorem for automata is quite helpful.

Lemma 7.1. *Let A be an automaton. Then, the following statements hold true:*

1. $L^d(A) = \bigcup_{B \in Fin(A)} L_{i_A, i_B}(A) L^d(B)$,
2. $L^{\lambda-d}(A) = \bigcup_{B \in Fin^\lambda(A)} L_{i_A, i_B}(A) L^{\lambda-d}(B)$,
3. $L^d(A) \subseteq L^{\lambda-d}(A)$.

7.2 Algebraic properties of disjunctive languages

We introduce the *(λ-)disjunctive closure*, M^d ($M^{\lambda-d}$), of a language $M \subseteq X^*$ by

$$M^d = \{\mathbf{x} \in M^\omega \mid \exists (w_i)_{i \in \mathbb{N}} : \{w_i \mid i \in \mathbb{N}\} = M^* \setminus \{\epsilon\} \wedge \mathbf{x} = w_1 \cdots w_i \cdots\},$$
$$M^{\lambda-d} = \{\mathbf{x} \in \operatorname{adh}(M^*) \mid M^* = \mathbf{T}(\mathbf{x})\}.$$

[21]We will prove later that ω–regular (λ)–disjunctive languages are exactly of the form $\bigcup_{1 \leq i \leq n} M_i {w_j}^\omega$ for some regular languages $M_i \subseteq X^*$ and words $w_i \in X^*$.

A sequence $\mathbf{x} \in X^\omega$ is in M^d if $\mathbf{x}$ is an infinite concatenation of words of M^* such that any word of M^* appears in this concatenation (which is equivalent to the fact that any word of M^* appears infinitely often). So, $\mathbf{x} \in X^\omega$ is in $M^{\lambda-d}$ if $\mathbf{x} \in \operatorname{adh}(M^*)$ and any word of M^* appears (infinitely often) as an infix of $\mathbf{x}$. Obviously, $M^d = (M^*)^d = M^* M^d$, $M^{\lambda-d} = (M^*)^{\lambda-d} = M^* M^{\lambda-d}$, $\mathbf{A}(M^d) = \mathbf{A}(M^{\lambda-d}) = \mathbf{A}(M^*)$.

Lemma 7.2. *Let A be an automaton. Then:*

1. *For any final sub-automaton B, $L^d(B) = (L_{i_B,i_B}(B))^d$.*
2. *For any λ-final sub-automaton B, $L^{\lambda-d}(B) = (L_{i_B,i_B}(B))^{\lambda-d}$.*

Proof. The inclusion $L^d(B) \subseteq (L_{i_B,i_B}(B))^d$ is obvious from the definitions. To prove "$\supseteq$" choose some $\mathbf{x} \in (L_{i_B,i_B}(B))^d$. Thus, there is an infinite sequence $(w_i)_{i\in\mathbb{N}}$ of words $w_i \in L_{i_B,i_B}(B) \setminus \{\epsilon\}$ with $\mathbf{x} = w_1 w_2 \cdots w_i \cdots$ where any word $u \in L_{i_B,i_B}(B) \setminus \{\epsilon\}$ must appear infinitely often as some w_i in $\mathbf{x}$. For any w_i we find a non–empty run r_i from i_B to i_B in B with $w_i = \lambda(r_i)$, thus, $\mathbf{r} = r_1 r_2 \cdots r_i \cdots$ is in $Run^\omega(B)$ with $\lambda(\mathbf{r}) = \mathbf{x}$. Let $C = \{c_1, c_2, \cdots\}$ be some enumeration of all circular runs in B from i_B to i_B. As B is strongly connected it is obvious that some run in $Run^\omega(B)$ is disjunctive if it contains C as sub–runs.

We define a new run $\mathbf{R} \in Run^\omega(B)$ inductively. First, put $R_1 = r_1 r_2 \cdots r_{i-1} c_1$ where $i = \mu y(\lambda(c_1) = \lambda(r_y))$, $t(R_1) = i$. Let R_n be defined such that R_n runs from i_B to i_B with $R_i \sqsubseteq R_j \sqsubseteq R_n$ for all $i \le j \le n$ and c_i is a sub–run of R_n for $1 \le i \le n$. For $i = \mu y(y > t(R_n) \wedge \lambda(c_{n+1}) = \lambda(r_y))$ define $R_{n+1} = R_n r_{t+1} \cdots r_{i-1} c_{n+1}$, where $t = t(R_n)$, and set $t(R_{n+1}) = i$. Thus, R_{n+1} runs from i_B to i_B, contains R_n as a prefix–run, and c_{n+1} as a sub–run. Thus, the unique run $\mathbf{R}$ in $\operatorname{adh}\{R_n | n \in \mathbb{N}\}$ contains all c_i as sub–runs and consequently is disjunctive with $\lambda(\mathbf{R}) = \lambda(\mathbf{r}) = \mathbf{x}$ i.e., $\mathbf{x} \in L^d(B)$. This proves the first part. To prove *2.* simply notice the following:[22]

$$\begin{aligned}
L^{\lambda-d}(B) &= \{\lambda(\mathbf{r}) | \mathbf{r} \in Run^\omega(B) \wedge \mathbf{r} \text{ is } \lambda - \text{disjunctive}\} \\
&= \{\lambda(\mathbf{r}) | \mathbf{r} \in \operatorname{adh} Run_{i_B,i_B}(B) \wedge \mathbf{r} \text{ is } \lambda - \text{disjunctive}\}, \\
&= \{\lambda(\mathbf{r}) | \mathbf{r} \in \operatorname{adh} Run_{i_B,i_B}(B) \wedge \forall r \in Run_{i_B,i_B}(B)\ \lambda(r) \in \mathbf{T}(\mathbf{r})\} \\
&= \{\mathbf{x} | \mathbf{x} \in \operatorname{adh} L_{i_B,i_B}(B) \wedge \forall u \in L_{i_B,i_B}(B)\ u \in \mathbf{T}(\mathbf{x})\} \\
&= \{\mathbf{x} \in \operatorname{adh}(L_{i_B,i_B}(B))^* | (L_{i_B,i_B}(B))^* \subseteq \mathbf{T}(\mathbf{x})\} \\
&= (L_{i_B,i_B}(B))^{\lambda-d}.
\end{aligned}$$

□

We may restate Lemma 7.1 as

Lemma 7.3. *Let A be an automaton. Then:*

1. $L^d(A) = \bigcup_{B \in Fin(A)} L_{i_A,i_B}(A) \left(L_{i_B,i_B}(B)\right)^d$,
2. $L^{\lambda-d}(A) = \bigcup_{B \in Fin^\lambda(A)} L_{i_A,i_B}(A) \left(L_{i_B,i_B}(B)\right)^{\lambda-d}$.

[22] Note that $Run_{i_B,i_B}(B)$ is the set of all finite runs in B from i_B to i_B.

We call an ω-language $F \subseteq X^\omega$ *d-regular* (λ-*d-regular*) if there are some non-empty regular languages $M_j, N_j \subseteq X^*$ with $F = \bigcup_{1\le j\le n} M_j {N_j}^d$ ($F = \bigcup_{1\le j\le n} M_j {N_j}^{\lambda-d}$, respectively), or if $F = \emptyset$.

Reg^d and $Reg^{\lambda-d}$ denote the classes of all d-regular and λ-d-regular languages.[23]

Theorem 7.4. $Rec^d = Reg^d$, $Rec^{\lambda-d} \subset Reg^{\lambda-d}$.

Proof. The relations $Rec^d \subseteq Reg^d$ and $Rec^{\lambda-d} \subseteq Reg^{\lambda-d}$ hold by Lemma 7.2 and the definition of (λ-) d-regularity. Obviously, $\{a,b\}^* a^\omega = \{a,b\}^*\{a\}^{\lambda-d} \in Reg^{\lambda-d}$, but any automaton A with $\{a,b\}^* a^\omega \subseteq L^{\lambda-d}(A)$ possesses also an infinite λ-disjunctive run $\mathbf{r}$ that stays in $\{a,b\}^*$, i.e. $D_{\{a,b\}} \subseteq L^{\lambda-d}(A)$ holds. Thus, $L^{\lambda-d}(A) \neq \{a,b\}^* a^\omega$ for any A (this is not true for disjunctivity, see example A_2), i.e., $Rec^{\lambda-d} \subset Reg^{\lambda-d}$ holds. It remains to prove $Reg^d \subseteq Rec^d$. For $F = \bigcup_{1\le j\le n} M_j {N_j}^d$, $N_j = {N_j}^*$, N_j, M_j regular, there exist automata C_j, B_j, sets $Q_j^f \subseteq Q_{C_j}$, such that

$$M_j = \bigcup_{q\in Q_j^f} L_{i_{C_j},q}(C_j),$$

$$N_j = L_{i_{B_j},i_{B_j}}(B_j),$$

where B_j may be chosen to be strongly connected and C_j to be trim. We construct an automaton A' with ϵ-arcs (i.e., arcs labeled with ϵ) by using $2j$ mutually disjoint versions of the automata C_j and B_j plus a new initial state i_A and placing ϵ-arcs from i_A to each i_{C_j} and from any $q \in Q_j^f$ to i_{B_j}. Finally, A follows from A' by a standard elimination of those ϵ-arcs (e.g., whenever (q,a,q') and (q',ϵ,q'') are edges, add a new edge (q,a,q''), keep (q,a,q'), and remove finally all ϵ-arcs). As C_j was chosen to be trim, then any disjunctive run in A from i_A that passes C_j must eventually reach B_j via some final state q in Q_j^f. Thus, $L^d(A) = F$, i.e. $Reg^d \subseteq Rec^d$. $\square$

Lemma 7.5. *The following statements hold true:*

1. $M \neq \emptyset \Rightarrow M^{\lambda-d} \neq \emptyset \neq M^d$,
2. $M^d \subseteq M^{\lambda-d}$,
3. $\exists M \subseteq X^*$: $M^{\lambda-d} \not\subseteq M^d$.

Proof. The first two statements are obvious. For *3.* we construct a language $M \subseteq X^*$ with $M = M^*$ but $M^{\lambda-d} \not\subseteq M^d$. For simplicity, we choose $M \subseteq \{a,b\}^*$. Let y_i be the i-th word in $\{a,b\}^*$ in lexicographic order, i.e., $|y_i| = \lfloor lg_2 i \rfloor$. Put $M = \{y_1 \dots y_i b^{2i+6} \mid i \in \mathbb{N}\}^*$, and let $\mathbf{y} = y_1 \dots y_i \dots$ be the infinite concatenation of all words of $\{a,b\}^*$ in their lexicographic order. Obviously, $\mathbf{y} \in \operatorname{adh} M$ and any $u \in M^* (= M)$ is an infix of some y_i, i.e. an infix of $\mathbf{y}$, thus $\mathbf{y} \in M^{\lambda-d}$. However, $\mathbf{y} \notin M^d$ as even $\mathbf{y} \notin M^\omega$ holds. This is easily seen as no word $v \in M$ is even a prefix of $\mathbf{y}$. Indeed, assume that $v \in M$ and $v \sqsubseteq \mathbf{y}$. As $v \in M$, v has $y_1 \dots y_i b^{2i+6}$ as a prefix for some i. But $\mathbf{y}$ has already $y_1 \dots y_i y_{i+1} y_{i+2}$

[23] Of course, d-regular languages are usually not ω-regular. The name "d-regular" only reflects the similarity in the definition of ω-regularity, and the fact that d-regularity is the algebraic counterpart of d-recognizability.

as a prefix and $|y_{i+1}y_{i+2}| \leq 2 \cdot lg_2(i+2) \leq 2i+6 = |b^{2i+6}|$. Thus, $y_{i+1}y_{i+2} \sqsubseteq b^{2i+6}$, a contradiction, as two consecutive words in lexicographic order must contain at least one letter a. □

We shall see later that the non-regularity of M, in the example used in the above proof, is essential as $M^d = M^{\lambda-d}$ holds for all regular languages M.

Lemma 7.6. *The following statements hold true:*

1. $M \subseteq N \wedge M^\omega \subseteq N^\omega \wedge \mathbf{A}(M) \subseteq \mathbf{A}(N) \wedge \mathbf{T}(M) \subseteq \mathbf{T}(N) \not\Rightarrow M^d \cap N^d \neq \emptyset \vee M^{\lambda-d} \cap N^{\lambda-d} \neq \emptyset$,
2. $M^d = N^d \wedge M^{\lambda-d} = N^{\lambda-d} \not\Rightarrow M \subseteq N \vee M^\omega \subseteq N^\omega$.

Proof. For *1.* let $M = a^*$ and $N = \{a,b\}^*$. They fulfill all hypotheses and $M^{\lambda-d} = M^d = a^\omega \notin N^{\lambda-d} = N^d = D_{\{a,b\}}$. To prove *2.* we choose $M = \{a,b\}^*$, $N = (a^*b)^*$. Thus, $M^{\lambda-d} = M^d = N^{\lambda-d} = N^d = D_{\{a,b\}}$, but $a^* \in M \setminus N$ and $a^\omega \in M^\omega \setminus N^\omega$. □

The above (trivial) examples warn us to be quite careful in finding relations between $M^{\lambda-d}$, M^d, M, and M^ω.

Lemma 7.7. *The following statements hold true:*

1. $M^d \subseteq N^d$ *or* $M^{\lambda-d} \subseteq N^{\lambda-d} \Rightarrow \mathbf{A}(M^*) \subseteq \mathbf{A}(N^*)$,
2. $M^{\lambda-d} \cap N^{\lambda-d} \neq \emptyset \Rightarrow \mathbf{T}(M^*) = \mathbf{T}(N^*)$.

Proof. *1.*: For $M^d \subseteq N^d$ we choose some $w \in \mathbf{A}(M)$. As $M^d = M^*M^d$ there is some z and $\mathbf{x}$ with $wz \in M^*$ and $wz\mathbf{x} \in M^d \subseteq N^d$. Thus $w \in \mathbf{A}(N^*)$. For $M^{\lambda-d} \subseteq N^{\lambda-d}$ use the same argument. To prove *2.* one simply notes that $w \in \mathbf{T}(M^*)$ and $\mathbf{x} \in M^{\lambda-d} \cap N^{\lambda-d}$ implies $w \in \mathbf{T}(\mathbf{x})$, thus $w \in \mathbf{T}(N^*)$. By symmetry, $\mathbf{T}(N^*) \subseteq \mathbf{T}(M^*)$ is also true. □

Lemma 7.8. *For every regular languages $M, N \subseteq X^*$, one has:*

$$\mathbf{S}(M^{\lambda-d}) = \mathbf{S}(N^{\lambda-d}) \Leftrightarrow \mathbf{T}(M^*) = \mathbf{T}(N^*).$$ [24]

Proof. For " $\Rightarrow$ " we note that $w \in \mathbf{T}(M^*)$ implies $z_1wz_2\mathbf{x} \in M^{\lambda-d}$, for some $z_1, z_2 \in M^*$ and $\mathbf{x} \in M^{\lambda-d} = M^*M^{\lambda-d}$, thus $wz_2\mathbf{x} \in \mathbf{S}(M^{\lambda-d}) = \mathbf{S}(N^{\lambda-d})$, i.e. $z_3wz_2\mathbf{x} \in N^{\lambda-d}$ for some $z_3 \in X^*$. Thus, $w \in \mathbf{T}(N^*)$. By symmetry one also concludes $T(N^*) \subseteq T(M^*)$.

For " $\Leftarrow$ " we note that $\mathbf{x} \in \mathbf{S}(M^{\lambda-d})$ implies $z\mathbf{x} \in M^{\lambda-d}$, for some $z \in \mathbf{A}(M^*)$. Thus, $(z\mathbf{x})[k] \in \mathbf{A}(M^*)$, $\forall k \in \mathbb{N}$. The relation $\mathbf{A}(M^*) \subseteq \mathbf{T}(M^*) = \mathbf{T}(N^*)$ holds true by assumption, i.e., $z\mathbf{x}[k] \in \mathbf{T}(N^*)$, $\forall k \in \mathbb{N}$. Thus, there is a sequence $(y_k)_{k\in\mathbb{N}}$ with $y_k \in \mathbf{A}(N^*)$ and $y_kz\mathbf{x}[k] \in \mathbf{A}(N^*) \forall k \in \mathbb{N}$. As N is regular N^* and $\mathbf{A}(N^*)$ are regular too; thus, the set of states $\mathbf{A}(N^*)/y_k = \{v \in X^* \mid y_kv \in \mathbf{A}(N^*)\}$ is finite. So, there is one $y \in X^*$ such that $yz\mathbf{x}[k] \in \mathbf{A}(N^*)$ for infinitely many k, i.e. for all $k \in \mathbb{N}$. Thus, $yz\mathbf{x} \in \operatorname{adh} \mathbf{A}(N^*)$. For $u \in N^*$ we conclude $u \in \mathbf{T}(N^*) = \mathbf{T}(M^*)$; thus, $u \in \mathbf{T}(\mathbf{x})$ as $\mathbf{x} \in M^{\lambda-d}$. But $yz\mathbf{x} \in \operatorname{adh} \mathbf{A}(N^*)$ and $\mathbf{T}(\mathbf{A}(N^*)) \subseteq \mathbf{T}(yz\mathbf{x})$ implies $yz\mathbf{x} \in N^{\lambda-d}$, thus, $\mathbf{x} \in \mathbf{S}(N^{\lambda-d})$. By symmetry one gets $\mathbf{S}(N^{\lambda-d}) \subseteq \mathbf{S}(M^{\lambda-d})$. □

[24] Remember, $\mathbf{S}(F)$ is the set of all suffixes of F.

Lemma 7.9. *The following statements are equivalent:*

1. $M^{\lambda-d} = N^{\lambda-d}$,
2. $\operatorname{adh} M^* = \operatorname{adh} N^*$,
3. $\mathbf{A}(M^*) = \mathbf{A}(N^*)$.

Proof. "*1.* ⇒ *2.*": From $\mathbf{x} \in \operatorname{adh} M^*$ we deduce $\mathbf{x}[k] \in \mathbf{A}(M^*) = \mathbf{A}(M^{\lambda-d}) = \mathbf{A}(N^{\lambda-d}) = \mathbf{A}(N^*)$ for all $k \in \mathbb{N}$; i.e., $\mathbf{x} \in \operatorname{adh} N^*$.

"*2.* ⇒ *3.*": As $w \in \mathbf{A}(M^*)$ implies $w\mathbf{x} \in M^{\lambda-d} = N^{\lambda-d}$ for some $\mathbf{x}$, one gets $w \in \mathbf{A}(N^*)$.

"*3.* ⇒ *1.*": Note that $\mathbf{x} \in M^{\lambda-d}$ implies $\mathbf{x} \in \operatorname{adh} M$ and $M^* \subseteq \mathbf{T}(\mathbf{x})$. $\mathbf{A}(M^*) = \mathbf{A}(N^*)$ implies immediately $\operatorname{adh} M^* = \operatorname{adh} N^*$. Further, $N^* \subseteq \mathbf{A}(N^*) = \mathbf{A}(M^*) \subseteq \mathbf{T}(\mathbf{x})$, i.e. $\mathbf{x} \in N^{\lambda-d}$. □

Combining Lemma 7.9 and Lemma 7.3 we get a decision procedure for testing whether $L^{\lambda-d}(A) = L^{\lambda-d}(B)$ for given automata A, B:

Theorem 7.10. *Equality in* $Rec^{\lambda-d}$ *is decidable.*

The same result holds for Rec^d, but this proof is far more involved.[25]

7.3 Reg^d and $Reg^{\lambda-d}$ are the same

If A is a deterministic automaton with some run r from some state q with label $\lambda(r) = w$ then this run is uniquely determined by q and w and will be denoted by $r_{q,w}$. We write $\delta(q, w) = q'$ if $r_{q,w}$ exists and ends in q'. Thus, δ refers always to deterministic automata.

Lemma 7.11. *For every deterministic automaton A one has: $L^d(A) = L^{\lambda-d}(A)$.*

Proof. It remains to prove the relation $L^{\lambda-d}(A) \subseteq L^d(A)$. We even can prove that any λ-disjunctive run in a deterministic automaton is also disjunctive. Therefore, we assume the existence of a λ-disjunctive but not disjunctive run $\mathbf{r}$ in A. I.e., there has to be some $q_0 \in Q^\omega(\mathbf{r})$ and some finite run r_0 in A from q_0 such that r_0 is no sub-run of $\mathbf{r}$. However, as $\mathbf{r}$ is λ-disjunctive, $u_0 = \lambda(r_0)$ is an infix of $\lambda(\mathbf{r})$ infinitely often. Thus, there is $q_1 \in Q^\omega(\mathbf{r})$ such that r_{q_1,u_0} is a sub-run of $\mathbf{r}$. As $q_0 \in Q^\omega(\mathbf{r})$, it follows that r_{q_1,u_0u_1} is a sub-run of $\mathbf{r}$ from q_1 to q_0 for some $u_1 \in X^*$. Thus, $r_{q_1,u_0u_1u_0}$ and $r_{q_0,u_0u_1u_0}$ are not sub-runs of $\mathbf{r}$ where $u_0u_1u_0$ is an infix of $\lambda(\mathbf{r})$ as $\mathbf{r}$ does not contain $r_0 = r_{q_0,u_0}$. By induction, we can find arbitrarily many states $q_0, q_1, q_2, \ldots, q_n \in Q^\omega(\mathbf{r})$ and words $u_0, u_1, \ldots, u_n \in X^*$ such that $r_{q_j}, u_0u_1u_0u_2u_0 \ldots u_0u_nu_0$ are not sub-runs of $\mathbf{r}$ for $1 \leq j \leq n$ but $u_0u_1u_0 \ldots u_0u_nu_0$ is an infix of $\lambda(\mathbf{r})$, a contradiction as $Q^\omega(\mathbf{r})$ is finite. □

[25] We shall give a rough idea of this proof in the following when we will look at further connections between disjunctivity and λ-disjunctivity.

The *deterministic closure*, A^{det}, of an arbitrary automaton $A = (Q_A, E_A, i_A)$ is defined as $A^{det} = (Q_A^{det}, E_A^{det}, \{i_A\})$, where $E_A^{det} \subseteq 2^{Q_A} \times X \times 2^{Q_A}$ is defined by $(S, a, S') \in E_A^{det}$ if $S' = \{q' \in S_A \,|\, \exists q \in S \, (q, a, q') \in E_A\}$. $Q_A^{det} \subseteq 2^{Q_A}$ is the set of all reachable states in A^{det}. Note that $(S, a, S') \in E_{A^{det}}$ does not mean that $(q, a, q') \in E_A$ holds for all $q \in S$ and all $q' \in S'$. Further, A being strongly connected does not imply that A^{det} is strongly connected. These facts lead to difficulties in gluing together parts of runs in A to form a larger run that corresponds to some given run in A^{det}. A gluing-technique is given in the following technical lemma.

Lemma 7.12. *In A^{det} the following properties hold:*

1. $S_1 \subseteq S_2 \Rightarrow \delta(S_1, w) \subseteq \delta(S_2, w)$.
2. $\delta(S, w) = S' \Leftrightarrow$
   - *i)* $\forall q \in S \ \exists q' \in S' \ \exists$ *run* r *in* A *from* q *to* q' *with* $\lambda(r) = w$*, and*
   - *ii)* $\forall q' \in S' \ \exists q \in S \ \exists$ *run* r *in* A *from* q *to* q' *with* $\lambda(r) = w$.
3. *If A is strongly connected with $E_A \neq \emptyset$ then for any disjunctive run $\mathbf{R}$ in A^{det} from $\{i_A\}$:*
   - *i)* $\exists S \in Q^\omega(\mathbf{R}) \ i_A \in S$
   - *ii)* $\forall S \in Q^\omega(\mathbf{R}) \ (i_A \in S \Rightarrow \exists S^\circ \in Q^\omega(\mathbf{R}) \ \exists v \in X^* \ (i_A \in S^\circ$ *and* $S^\circ = \delta(S, v) = \delta(\{i_A\}, v)))$.

Proof. The first two statements are obvious. For *3.*, note that $\mathbf{R}$ has to reach some final sub-automaton B^{det} of A^{det}; thus, $\mathbf{R} = \mathbf{R}[\ell]\mathbf{R}'$ for some ℓ where $\mathbf{R}'$ runs completely inside B^{det}. For any $S' \in Q^\omega(\mathbf{R}) = Q_B^{det}$ there is some k $(> \ell)$ and $S' = \delta(\{i_A\}, u)$ with $u = \lambda(\mathbf{R}[k])$. By Lemma 7.12, *2* there is a run r in A from i_A to q' with $\lambda(r) = u$ and $q' \in S'$. As A is strongly connected there is some run r' in A from q' to i_A. Thus,

$$i_A \in \delta(\{i_A\}, u\lambda(r')) \in Q_B^{det} = Q^\omega(\mathbf{R}),$$

which proves *3.i)*.

Now, let S be some state in $Q^\omega(\mathbf{R})$ $(= Q_B^{det})$ with $i_A \in S$. $S = \delta(\{i_A\}, u)$ with $u = \lambda(\mathbf{R}[k])$ for some k. By Lemma 7.12, *2.ii)* we know the existence of some run r in A from i_A to i_A with $\lambda(r) = u$. We define $S_i = \delta(\{i_A\}, u^i)$ for $i \in \mathbb{N}$, i.e., $S = S_1$. By Lemma 7.12.*1* we conclude from $\{i_A\} \subseteq S$ that $\delta(\{i_A\}, u) \subseteq \delta(S, u)$ holds, thus $S_1 \subseteq S_2$ as $S_1 = \delta(\{i_A\}, u)$ and $S_2 = \delta(\{i_A\}, u^2) = \delta(S_1, u)$. By induction, one concludes $\{i_A\} \subseteq S_1 \subseteq S_2 \subseteq \ldots$. As $Q^\omega(\mathbf{R})$ is finite this sequence must become stationary with some S_j $(= S_{j+1})$. For $S^\circ = S_j$ the desired properties $i_A \in S^\circ$ and $S^\circ = \delta(\{i_A\}, v) = \delta(S, v)$ hold with $v = u^j$. □

Lemma 7.13. *If A is strongly connected, then $L^d(A^{det}) \subseteq L^d(A)$.*

Proof. For $\mathbf{x} \in L^d(A^{det})$ there is a disjunctive run $\mathbf{R}$ in A^{det} from $\{i_A\}$ with $\lambda(\mathbf{R}) = \mathbf{x}$. Let $C = \{c_1, c_2, \ldots\}$ be an enumeration of all finite runs in A from i_A to i_A. We construct a disjunctive run $\mathbf{r}$ with $\mathbf{x} = \lambda(\mathbf{r})$ as follows.

By Lemma 7.12, *3.i)* there is some state S in $Q^\omega(\mathbf{R})$ with $i_A \in S$. Further, $S = \delta(\{i_A\}, \lambda(\mathbf{R}[k_0]))$ for some $k_0 \in \mathbb{N}$. By Lemma 7.12, *2.ii)* we find a run r_0 in A from i_A to i_A with $\lambda(r_0) = \lambda(\mathbf{R}[k_0])$.

Assume we have already defined r_n such that the following holds:

- r_n is a run in A from i_A to i_A,
- c_j is a sub-run of r_n for $1 \leq j \leq n$,
- $\mathbf{R}[k_n]$ runs in A^{det} from $\{i_A\}$ to some state $S_n \in Q^\omega(\mathbf{R})$ with $i_A \in S_n$, for $k_n = |r_n|$,
- $\lambda(r_n) = \lambda(\mathbf{R}[k_n])$.

By Lemma 7.12, *3* we can find a word v and a run $r_{S_n,v}$ from S_n to some $S^\circ \in Q^\omega(\mathbf{R})$ with $i_A \in S^\circ$ and $\delta(S_n, v) = \delta(\{i_A\}, v) = S^\circ$. Note that c_{n+1} is the $(n+1)$st finite run in A from i_A to i_A and $i_A \in S^\circ \in Q^\omega(\mathbf{R})$ holds. Thus, $\mathbf{R}$ may use $r_{S^\circ,w_{n+1}}$ with $w_{n+1} = \lambda(c_{n+1})$ infinitely often and has to do so: there is some finite run R in A^{det} from S° to S° such that $\mathbf{R}[k_n]r_{S_n,v}Rr_{S^\circ,w_{n+1}}$ is a prefix of $\mathbf{R}$. From $\delta(S_n, v) = \delta(\{i_A\}, v)$ we know $\delta(S_n, v\lambda(R)) = \delta(\{i_A\}, v\lambda(R)) = S^\circ \ni i_A$. By Lemma 7.12, *2.ii)* there is thus a run r' in A from i_A to i_A with $\lambda(r') = v\lambda(R)$. We now define $r_{n+1} = r_n r' c_{n+1}$ and conclude:

- r_{n+1} is a run in A from i_A to i_A,
- c_j is a sub-run of r_{n+1} for $1 \leq j \leq n+1$,
- for $k_{n+1} = |r_{n+1}|$, $\mathbf{R}[k_{n+1}]$ $(= \mathbf{R}[k_n]r_{S_n,v}Rr_{S^\circ,w_{n+1}})$ runs in A^{det} from $\{i_A\}$ to some state $S_{n+1} \in Q^\omega(\mathbf{R})$ with $i_A \in S_{n+1}$,
- $\lambda(r_{n+1}) = \lambda(\mathbf{R}[k_{n+1}])$.

As $r_n \sqsubseteq r_{n+1}, |r_n| < |r_{n+1}|$ there is a run $\mathbf{r} \in \operatorname{adh}\{r_n \mid n \in \mathbb{N}\}$ with all required properties. □

Lemma 7.14. *If A strongly is connected, then $L^{\lambda-d}(A) \subseteq L^{\lambda-d}(A^{det})$.*

Proof. For $\mathbf{x} \in L^{\lambda-d}(A)$ there is a λ-disjunctive run $\mathbf{r}$ in A from i_A with $\lambda(\mathbf{r}) = \mathbf{x}$. In view of the construction of A^{det} there is exactly one infinite run $\mathbf{R}$ in A^{det} from $\{i_A\}$ with $\lambda(\mathbf{R}) = \lambda(\mathbf{r}) = \mathbf{x}$. For $S \in Q^\omega(\mathbf{R})$ and any finite run R in A^{det} from S with $v = \lambda(R)$ we find a finite run r in A from some $q \in S$ with $\lambda(r) = v$. By λ-disjunctivity of $\mathbf{r}$ we conclude that $v \in \lambda(\mathbf{r})$ holds, thus $v \in \lambda(\mathbf{R})$ and $\mathbf{R}$ is also λ-disjunctive. □

Now we have completed all necessary steps to conclude the main results.

Theorem 7.15. *If A is strongly connected, then $L^d(A) = L^{\lambda-d}(A)$.*

Proof. Use, in order, Lemma 7.1, Lemma 7.14, Lemma 7.11, and Lemma 7.13 to get:

$$L^d(A) \subseteq L^{\lambda-d}(A) \subseteq L^{\lambda-d}(A^{det}) = L^d(A^{det}) \subseteq L^d(A).$$

□

Theorem 7.16. *If $M \in Reg$, then $M^d = M^{\lambda-d}$.*

Proof. As $M^d = (M^*)^d$, $M^{\lambda-d} = (M^*)^{\lambda-d}$ we may assume that $M = M^*$ holds. Thus, there is a strongly connected automaton A with $M = L_{i_A,i_A}(A)$. We conclude by Lemma 7.2 and Theorem 7.15:

$$M^d = (L_{i_A,i_A}(A))^d = L^d(A) = L^{\lambda-d}(A) = (L_{i_A,i_A}(A))^{\lambda-d} = M^{\lambda-d}.$$

□

Theorem 7.17. *$Rec^{\lambda-d} \subset Reg^{\lambda-d} = Reg^d = Rec^d$.*

Proof. Theorem 7.16 implies $Reg^{\lambda-d} = Reg^d$ while Theorem 7.4 implies the rest. □

We call an automaton A *finally deterministic* if any final sub-automaton of A is deterministic. In Lemma 7.3 the final sub-automata B are strongly connected. We thus may replace them by B^{det}, following the proof of Theorem 7.15. As an obvious consequence we deduce that any disjunctive language $L \subseteq X^*$ is the disjunctive language $L^d(A)$ of some finally deterministic automaton.

Theorem 7.18. *Equality in Rec^d is decidable.*

Sketch of proof. This proof requires all facts on disjunctive runs in A and A^{det} but, nevertheless, it is not an easy consequence of them. We only present a sketch of the proof (from the Ph.D. Thesis of Rehrmann [39]). We call a relation $\tau \subseteq Q_A \times Q_A$ *weakly continuable* if τ is symmetric and $q\tau q'$ implies the following:

$\forall q_1 \in Q_A$ $\forall$ runs r in A from q to q_1 $\exists q_2, q_2' \in Q_A$ $\exists$ run r_1 in A from q_1 to q_2
$\exists$ run r' in A from q' to q_2' $(\lambda(rr_1) = \lambda(r') \ \wedge \ q_2 \tau q_2')$.

We define $q \sim_{weak} q'$ if there exists a weak continuable relation τ with $q\tau q'$. For $A = (Q_A, E_A, i_A)$, $q \in Q_A$, define $A_q = (Q_A, E_A, q)$. Define $q \sim_d q'$ if $L^d(A_q) = L^d(A_{q'})$. Then the following important connection between $\sim_{weak}$ and $\sim_d$ holds true for finally deterministic automata:

$$q \sim_{weak} q' \Longleftrightarrow q \sim_d q',$$ [26]

for all $q, q' \in Q_A$. As $\sim_{weak}$ turns out to be decidable, one has a decision procedure for the equality-problem in Rec^d. □

[26] As $\sim_{weak}$ is a variant of bisimulation, this equation presents an interesting connection between disjunctivity and bisimulations. Bisimulations play an important role in the theory of concurrent computations.

7.4 Topological properties of disjunctive languages

Here we study closure properties of Reg^d with respect to finite and countable union and intersection, and compute the complexity of Reg^d in the Borel hierarchy. First, we answer the simple question whether Reg^d may contain any ω–regular, closed or open languages.

We need the following simple technical lemma.

Lemma 7.19. *Let B be a strongly connected and deterministic automaton. If $L^d(B)$ possesses an ultimately period word uw^ω for some u, $w \in X^*$, then $L^d(B) = w^\omega$.*

Proof. As B is deterministic we identify here runs from i_B with their labeling words. It suffices to show that there are no two runs r_1, r_2 in B from i_B with $r_1 \not\sqsubseteq r_2$ and $r_2 \not\sqsubseteq r_1$ whenever $L^d(B)$ possess at least one ultimately periodic word. Thus, we assume the existence of a word w and two different letters a, b such that wa and wb run from i_B. Both runs can be continued to reach i_B again. Thus, there are cyclic runs $c_1 = wac'_1$ and $c_2 = wbc'_2$ from i_B to i_B. Define $C_1 \colon = c_1^{|c_2|}$, $C_2 \colon = c_2^{|c_1|}$. Thus, $|C_1| = |C_2|$ and the infinite run

$$\mathbf{C} := C_1 C_2 C_1^2 C_2^2 \cdots C_1^i C_2^i \cdots$$

is not ultimately periodic. However, if there is a disjunctive ultimately periodic run $\mathbf{r} = uw^\omega$ in B, then any arbitrarily large part of $\mathbf{C}$ must be a sub–run of $\mathbf{r}$, a contradiction, as $\mathbf{r}(|w| \cdot |C_1| \cdot n + k)$ must always be the same letter for n large enough in $\mathbf{r}$, for any k, but has to be a and b in $\mathbf{C}$, for some k and infinitely many values n. □

Theorem 7.20. *The following statements are true:*

1. $L \in Reg^d \cap Reg^\omega \Leftrightarrow L = \bigcup_{1 \le i \le n} M_i w_i^\omega$ *for some regular* $M_i \subseteq X^*$ *and* $w_i \in X^*$.
2. $L \in Reg^d \cap \mathbf{G} \Leftrightarrow L = \emptyset$.
3. $L \in Reg^d \cap \mathbf{F} \Leftrightarrow L = L^d(A)$ *for some automaton* A *with* $\lambda(Run^\omega(A)) = L^d(A)$.
4. $Reg^d \cap \mathbf{F} \subseteq Reg^\omega$.

Proof. The first statement is an obvious consequence of Lemma 7.19. For $L \in Reg^d$ there is a finally deterministic automaton A with $L = L^d(A) = \bigcup_{1 \le i \le m} M_i L^d(B_i)$ for some regular languages M_i and final deterministic automata B_i. Thus, $L^d(B_i) = w_i^\omega$ for some $w_i \in X^*$ or $L^d(B_i)$ possesses no ultimately periodic word, i.e., $L = L_1 \cup L_2$ with an ω–regular language $L_1 = \bigcup_{1 \le i \le n} M_i w_i^\omega$ and some language L_2 without any ultimately period word. If L is in addition ω–regular the same must hold for $L_2 = L - L_1$, thus $L_2 = \emptyset$ as any non–empty ω–regular language must possess an ultimately periodic word.

2.: L is open if $L = MX^\omega$, for some $M \subseteq X^*$. As we assume that $|X| > 1$ holds $\emptyset \neq L$ must possess infinitely many ultimately periodic words in contrast to disjunctive languages.

3.: Let $L = L^d(A)$ where without any loss of generality any state q of A possesses a run from q into some final sub–automaton of A. Thus, any finite run in A from i_A can be continued into some disjunctive run, which implies $\operatorname{adh}(L^d(A)) = \lambda(Run^\omega(A))$. As $L = \operatorname{adh} L$ for any closed ω–language *3.* follows. As $\lambda(Run^\omega(A))$ is ω–regular *4.* follows as well. □

Lemma 7.21. *Reg^d is not closed under countable union or countable intersection.*

Proof. Clearly, $a^{j^2}b^\omega \in Reg^d$ for any $j \in \mathbb{N}$, but $\bigcup_{j>0} a^{j^2}b^\omega = \{a^{j^2} \mid j \in \mathbb{N}\}b^\omega$ is not in Reg^d. This is easily seen as $L^d(A) = \{a^{j^2} \mid j \in \mathbf{N}\}b^\omega$ implies that all final sub-automata of A possess only b–arcs. Thus, if we drop all b-arcs in A we get an automaton A' that would recognize $\{a^{j^2} \mid j \in \mathbb{N}\}$, a contradiction as $\{a^{j^2} \mid j \in \mathbb{N}\}$ is not regular. Also, $L_k = \{baba^2ba^3 \dots ba^k\}\{a,b\}^\omega \in Reg^d$, but $L = \bigcap_{k>1} L_k = \{baba^2ba^3 \dots ba^k \dots\} \notin Rec^d$. This is easily seen as a disjunctive language consisting of only one infinite word must be of the form uw^ω, following the proof of Lemma 7.19. □

As a consequence, *there is no topology for X^ω such that Reg^d is the class of open or of closed sets in this topology.*

Lemma 7.22. *The class Reg^d is closed under finite union and intersection.*

Proof. For $M_i = L^d(A_i)$ one easily constructs a new (non-deterministic) automaton A (that consists mainly of two disjoint versions of A_1 and A_2 plus one new initial state from which one may enter A_1 or A_2) such that $L^d(A) = L^d(A_1) \cup L^d(A_2)$ holds. However, a construction of an automaton A with $L^d(A) = L^d(A_1) \cap L^d(A_2)$ is rather involved. Two states $q_1, q_2 \in Q_A$ are *equivalent* if

$$\bigcup_{q \in Q_A} L_{q_1,q}(A) = \bigcup_{q \in Q_A} L_{q_2,q}(A).$$

An automaton is *reduced* if no two different states are equivalent. It turns out that A may always be chosen to possess only deterministic and reduced final sub–automata. From switching theory the concept of a *homing-sequence* is adopted: a word $w \in X^*$ is a homing-sequence if all finite runs in A with labeling w end in the same state. It turns out that strongly connected, deterministic, reduced automata always possess a homing sequence. As disjunctive runs have to use such a homing sequence we have some control on disjunctive runs r_i in A_i with $\lambda(r_1) = \lambda(r_2)$: both have to reach isomorphic sub-automata where they become eventually "synchronized". Exploiting those facts one may construct an automaton A with $L^d(A) = L^d(A_1) \cap L^d(A_2)$. Details may be found in Nolte [31]. □

We have omitted some details in the previous proof as Lemma 7.22 is not needed in the following in the study of the complexity of Reg^d in the Borel hierarchy. It is a well-known result that $Reg^\omega \subseteq \mathbf{F}_{\sigma\delta} \cap \mathbf{G}_{\delta\sigma}$ holds in the Borel hierarchy.[27]

Theorem 7.23. *$Reg^d \subseteq \mathbf{F}_{\sigma\delta} \cap \mathbf{G}_{\delta\sigma}$.*

[27] Originally, it was one hope to find a natural acceptance concept–via disjunctivity–for Reg^d to lie above Reg^ω.

Proof. For $M \subseteq X^*$ we compute the complexity of M^δ, $\operatorname{adh} M$, and M^d.

$$M^\delta = \{\mathbf{x} \in X^\omega \mid \operatorname{card}(\mathbf{A}(\mathbf{x}) \cap M) = \infty\} = \bigcap_{i \in \mathbb{N}} \left(M \cap \bigcup_{j \geq i} X^j \right) X^\omega \in \mathbf{G}_\delta.$$

Analogously, $\operatorname{adh} M = \mathbf{A}(M)^\delta \in \mathbf{G}_\delta$. For $u \in X^*$ we get

$$\{u\}M^\delta = \bigcap_{i \in \mathbb{N}} \{u\} \left(M \cap \bigcup_{j \geq i} X^j \right) X^\omega \in \mathbf{G}_\delta.$$

For a regular M,

$$M^d = M^{\lambda-d} = \operatorname{adh} M^* \cap \bigcap_{u \in M} (X^* u)\, X^\omega \in \mathbf{G}_\delta,$$

and

$$\{u\}M^d = \{u\}M^{\lambda-d} \in \mathbf{G}_\delta.$$

Thus, for any automaton A we get

$$\begin{aligned} L^d(A) &= \bigcup_{B \in Fin^d(A)} L_{i_A, i_B}(A)(L_{i_B, i_B}(B))^d \\ &= \bigcup_{B \in Fin^d(A)} \left(\bigcup_{u \in L_{i_A, i_B}(B)} \{u\} \left(L_{i_B, i_B}(B)\right)^d \right) \in \mathbf{G}_{\delta\sigma}, \end{aligned}$$

showing that $Reg^d = Rec^d \subseteq \mathbf{G}_{\delta\sigma}$.

To prove $Reg^d \subseteq F_{\sigma\delta}$ a concept of k-disjunctivity in quite helpful. An infinite run $\mathbf{r}$ in A is *k-disjunctive* if for all $q \in Q^\omega(\mathbf{r})$ and for all runs r in A from q of length k it is true that r is a sub-run of $\mathbf{r}$. A *k-disjunctive language* is defined by $L^{k-d}(A) = \{\mathbf{x} \in X^\omega \mid \exists \mathbf{r} \in Run^\omega(A)\ (\mathbf{r} \text{ is } k\text{-disjunctive} \wedge \mathbf{x} = \lambda(\mathbf{r}))\}$. However, as there are only finitely many runs of a fixed length k from any state $q \in Q_A$ one may describe them by a finite formula. In fact, it is a rather simple exercise to describe $L^{k-d}(A)$ by a S1S-formula of "second order logic of one successor". By a famous result of Büchi [2] $L^{k-d}(A)$ is thus ω-regular for all $k \in \mathbb{N}$. The relation $Reg^\omega \subseteq \mathbf{F}_{\sigma\delta}$ implies $L^{k-d}(A) \in \mathbf{F}_{\sigma\delta}$, for all $k \in \mathbb{N}$ and one easily sees that $L^d(A) = \bigcap_{k>0} L^{k-d}(A)$. As $\mathbf{F}_{\sigma\delta}$ is closed under countable intersection this implies $L^d(A) \in \mathbf{F}_{\sigma\delta}$, thus $Rec^d \subseteq \mathbf{F}_{\sigma\delta}$. □

Disjunctivity, called *path-fairness*, and λ-disjunctivity, called *word-fairness*, have been intensively researched by D. Nolte, L. Priese, R. Rehrmann and U. Willecke-Klemme in the area of fairness. There, Rec^d, $Rec^{\lambda-d}$, Reg^d, and $Reg^{\lambda-d}$ were introduced by Priese [36]. A hierarchy of variants of k-disjunctive languages was found in the Ph.D. thesis of U. Willecke-Klemme, see also Priese and Willecke-Klemme [37]. Some topological properties of disjunctive languages are found in the Masters Thesis of Nolte [31] and some connections to ultra-metrics were studied by Darondeau, Nolte, Priese, and Yoccoz, see [14], [34].

Many of the relations between λ-disjunctive and disjunctive languages are by Priese and Rehrmann [35]. The decidability of equality in Rec^d is by Rehrmann [39].[28]

[28] In these papers the authors operate with a slightly different concept for automata, namely with non-deterministic finite automata with ϵ-edges. As ϵ-edges seem to be not very natural in the study of disjunctive runs we had to change the concept and thus several proofs.

Acknowledgement

In writing the final version of this paper we benefitted from Peter Hertling's critical remarks.

References

[1] É. Borel, Les probabilités dénombrables et leurs applications arithmétiques, *Rend. Circ. Mat. Palermo* **27** (1909), 247–271.

[2] J.R. Büchi, Weak second-order arithmetic and finite automata, *Zeit. Math. Logik Grund. d. Math. 6* (1960), 66–92.

[3] J.R. Büchi, On a decision method in restricted second-order arithmetic, *Proc. of the Int. Cong. on Logic, Methodology, and Philosophy of Sciences* (1960), Stanford University Press, California.

[4] C. Calude, *Theories of Computational Complexity,* North-Holland, Amsterdam, 1988.

[5] C. Calude, Borel normality and algorithmic randomness. In: G. Rozenberg, A. Salomaa (eds.), *Developments in Language Theory*, World Scientific, Singapore, 1994, 113–129.

[6] C. Calude, *Information and Randomness: An Algorithmic Perspective,* Springer-Verlag, Berlin, 1994.

[7] C. Calude, H. Jürgensen, Randomness as an invariant for number representations. In: H. Maurer, J. Karhumäki, G. Rozenberg, (eds.). *Results and Trends in Theoretical Computer Science,* Springer-Verlag, Berlin, 1994, 44–66.

[8] C. Calude, T. Zamfirescu, Most Numbers Obey No Probability Laws, *Technical Report No 112, 1995*, Department of Computer Science, The University of Auckland, New Zealand, 4 pp.

[9] C. Calude, T. Zamfirescu, The typical number is a lexicon, *NZ J. Math.* (in press).

[10] C. Calude, S. Yu. Language-theoretic complexity of disjunctive sequences, in M. E. Houle, P. Eades (eds.), *Proceedings CATS'96 (Computing: the Australian Theory Seminar)*, Melbourne, Australia, 29-39 January 1996, 175-179.

[11] G. J. Chaitin, *Information, Randomness and Incompleteness, Papers on Algorithmic Information Theory,* World Scientific, Singapore, 2nd ed., 1990.

[12] D. G. Champernowne, The construction of decimals normal in the scale of ten, *J. London Math. Soc.* **8** (1933), 254–260.

[13] A. H. Copeland, P. Erdös, Note on normal numbers, *Bull. Amer. Math. Soc.* **52** (1946), 857–860.

[14] P. Darondeau, D. Nolte, L. Priese, S. Yoccoz, Fairness, distances, and degrees, *Theoret. Comput. Sci.* **97** (1992), 131-142

[15] E. Dolzhenko, Boundary properties of arbitrary functions, *Izv. Akad. Nauk SSSR Ser. Mat.* **31** (1967), 3–14.

[16] S. I. El-Zanati, W. R. R. Transue, On dynamics of certain Cantor sets, *J. Number Theory*, **36** (1990), 246–253.

[17] G. H. Hardy, E. M. Wright, *An Introduction to the Theory of Numbers*, Clarendon Press, Oxford, 5th ed., 1979.

[18] P. Hertling. Disjunctive ω-words and real numbers, *Technical Report* 180, Fernuniversität Hagen, 1995.

[19] P. Hertling. Disjunctive ω-words and real numbers, *J. UCS* 2 (1996), 549-568.

[20] G. Istrate, Gh. Păun, Some combinatorial properties of self-reading sequences, *Discrete Appl. Math.* **55** (1994), 83–86.

[21] H. Jürgensen, G. Thierrin, Some structural properties of ω-languages, *13th Nat. School with Internat. Participation "Applications of Mathematics in Technology"*, Sofia, 1988, 56–63.

[22] S. C. Kleene, Representation of events in nerve nets and finite automata, *Automata Studies*, Princeton Univ. Press, 1956, 3–42.

[23] W. Kuich, On the entropy of context-free languages, *Inform. Control* **16** (2), (1970), 173–200.

[24] L. Kuipers, H. Niederreiter, *Uniform Distribution of Sequences*, John-Wiley & Sons, New York, 1974.

[25] K. Kuratowski, *Topology* I, Academic Press, New York, 1966.

[26] R. Lindner, L. Staiger, *Algebraische Codierungstheorie – Theorie der sequentiellen Codierungen*, Akademie-Verlag, Berlin, 1977.

[27] P. Martin-Löf, The definition of random sequences, *Inform. and Control* **9** (1966), 602–619.

[28] R. McNaughton, Testing and generating infinite sequences by a finite automaton, *Information and Control* **9** (1966), 521–530.

[29] W. Merzenich, L. Staiger, Fractals, dimension, and formal languages, *RAIRO–Inform. Théor.* **28** (3–4) (1994), 361–386.

[30] D. E. Muller, Infinite sequences and finite machines, *Switching Circuit Theory and Logical Design, Proc. Fourth Ann. Symp.*, Inst. of Electrical and Electronic Engineers, Chicago, 1963.

[31] D. Nolte, *Eigenschaften pfadfairer Sprachen hinsichtlich Topologie und Reduzierbarkeit,* Master Thesis, Dept. Comp. Sci., Univ. Paderborn, 1989.

[32] J. C. Oxtoby, *Measure and Category*, Springer-Verlag, Berlin, 1971.

[33] J. C. Oxtoby, S. M. Ulam, Measure-preserving homeomorphisms and metrical transitivity, *Ann. Math.* **42** (1941), 874–925.

[34] L. Priese, Approaching fair computations by ultra metrics. In: I. Guessarian (ed.), *Semantics of Systems of Concurrent Programs, Proceedings,* Lecture Notes in Comput. Sci. **469** (1990), 420–433.

[35] L. Priese, R. Rehrmann, Deciding the path and word–fair equivalence problem. In: M. Joseph (ed.), *Formal Techniques in Real-Time and Fault-Tolerant Systems,* Lecture Notes in Comput. Sci. **329** (1988), 200–222.

[36] L. Priese, R. Rehrmann, U. Willecke–Klemme, An introduction to the regular theory of fairness, *Theoret. Comput. Sci.* **54** (1987), 139–163.

[37] L. Priese, U. Willecke-Klemme, *Fair Languages: A Study of Hierarchical Structures,* TR 75, Dept. Comp. Sci., Univ. Paderborn, 1990.

[38] W. M. Priestley, Sets thick and thin, *Amer. Math. Monthly* **83** (1976), 648–650.

[39] R. Rehrmann, *Pfad- und Wort-Fairness als Akzeptanzbedingung einer Klasse nicht-ω-regulärer, aber erkennbarer Sprachen,* Ph.D. Thesis, Dept. Comp. Sci., Univ. Paderborn, 1988.

[40] W. M. Schmidt, On normal numbers, *Pacific J. Math.*, **10** (1960), 661–672.

[41] L. Staiger, Measure and category in X^ω. In: J. Flachsmeyer, Z. Frolik and F. Terpe (eds.), *Topology and Measure II,* Part 2, Wiss. Beitr. Ernst–Moritz–Arndt–Univ. Greifswald, 1980, 129–136.

[42] L. Staiger, Finite-state ω-languages, *J. Comput. System. Sci.* **27** (1983), 434–448.

[43] L. Staiger, The entropy of finite-state ω-languages, *Problems Control Inform. Theory* **14** (5) (1985), 383–392.

[44] L. Staiger, Hierarchies of recursive ω-languages. J. Inform. Process. Cybernetics EIK 22 (1986) 5/6, 219–241.

[45] L. Staiger, Sequential mappings of ω-languages. RAIRO Infor. théor. et Appl. 21 (1987) 2, 147–173.

[46] L. Staiger, Kolmogorov complexity and Hausdorff dimension, *Inform. and Comput.* **103** (2) (1993), 159–194. Preliminary version In: J. Csirik, J. Demetrovics (eds.), *Proc. Intern. Conf. "Fundamentals of Computation Theory"*, Lecture Notes in Comput. Sci. **No. 380**, Springer-Verlag, Berlin, 1989, 334–343.

[47] L. Staiger, On ω-power languages, in: New Trends in Formal Languages, Control, Cooperation, and Combinatorics, (Gh. Păun and A. Salomaa Eds.), Lecture Notes in Comput. Sci. No. 1218, Springer–Verlag, Berlin 1997. pp. 377 – 393.

[48] L. Staiger, K. Wagner, Automatentheoretische und automatenfreie Charakterisierungen topologischer Klassen regulärer Folgenmengen, *Elektron. Informationsverarb. Kybernet. EIK*, **10** (1974), 379–392.

[49] A. Szilard, S. Yu, K. Zhang, J. Shallit, Characterizing regular languages with polynomial densities. In: I. M. Havel, V. Koubek (eds.), *Proceedings 17th International Symposium Mathematical Foundations of Computer Science*, Lecture Notes in Comput. Sci. 629, Springer-Verlag, Berlin, 1992, 494–503.

[50] W. Thomas, Automata on infinite objects, in J. Van Leeuwen (Ed.), *Handbook of Theoretical Computer Science*, Vol. B, 133–191, Elsevier, Amsterdam, 1990.

[51] R. von Mises, *Probability, Statistics and Truth*, G.Allen and Unwin, London; Macmillan, New York, 2nd Revised English Edition Prepared by Hilda Geiringer, 1961.

[52] L. Zajíček, Porosity and σ-porosity, *Real Analysis Exchange* **13** (1987/88), 314–350.

[53] T. Zamfirescu, Porosity in convexity, *Real Analysis Exchange* **15** (1989/90), 424–436.