Notre Dame Journal of Formal Logic
Volume 57, Number 3, 2016

Some Remarks on Real Numbers Induced by First-Order Spectra

Sune Kristian Jakobsen and Jakob Grue Simonsen

Abstract The spectrum of a first-order sentence is the set of natural numbers occurring as the cardinalities of finite models of the sentence. In a recent survey, Durand et al. introduce a new class of real numbers, the spectral reals, induced by spectra and pose two open problems associated to this class. In the present note, we answer these open problems as well as other open problems from an earlier, unpublished version of the survey.

Specifically, we prove that (i) every algebraic real is spectral, (ii) every automatic real is spectral, (iii) the subword density of a spectral real is either 0 or 1, and both may occur, and (iv) every right-computable real number between 0 and 1 occurs as the subword entropy of a spectral real.

In addition, Durand et al. note that the set of spectral reals is not closed under addition or multiplication. We extend this result by showing that the class of spectral reals is not closed under any computable operation satisfying some mild conditions.

1 Spectral Reals

We assume familiarity with basic first-order logic and its model theory at the level of standard introductory texts such as Ebbinghaus, Flum, and Thomas [9] and Enderton [10]. Recall that the *spectrum* of a sentence ϕ in first-order logic is the set of nonnegative integers n such that ϕ has a model of cardinality n.

In their survey [7], Durand et al. introduce a new class of real numbers induced by spectra.

Received May 19, 2011; accepted October 9, 2013
First published online March 24, 2016
2010 Mathematics Subject Classification: Primary 03C13, 68Q15, 68Q19; Secondary 11U05, 11U09
Keywords: first-order logic, spectral reals, computability theory, computational complexity
 10.1215/00294527-3489987

Definition 1.1 For a first-order sentence ϕ, let $\chi_\phi(n)$ be the characteristic function of the spectrum of ϕ. For $a \in \mathbb{Z}$, the *real number induced by ϕ and a* is $r_{\phi,a} = a + \sum_n \chi_\phi(n)2^{-n}$.

A real number r is said to be (first-order) *spectral* if $r = r_{\phi,a}$ for some first-order sentence ϕ and $a \in \mathbb{Z}$.

Open Question 11 in [7] asks (i) whether there are any irrational algebraic reals that are also spectral, and (ii) whether all automatic reals are spectral. In addition, an earlier version of the survey by Durand et al. [8, Open Question 14] asked whether the subword density of all spectral numbers is strictly less than 1 (or even 0).

As it turns out, one of the classic characterizations of the set of spectra furnishes straightforward answers to all of the above questions. Specifically, the *Jones–Selman characterization* in [15] states that *if* $\mathbf{b}(n)$ *is the standard representation of natural number n in binary, then $A \subseteq \mathbb{N}$ is a spectrum if and only if* $\{\mathbf{b}(n) : n \in A\} \in$ NE, where NE is the set of sets $B \subseteq \{0,1\}^*$ with the property that there exists a nondeterministic multitape Turing machine that decides B and on input x terminates in time $2^{O(|x|)}$. Thus A is a spectrum if and only if there is a nondeterministic Turing machine that decides if $n \in A$ in time $2^{(O(|\mathbf{b}(n)|))} = n^{O(1)}$. Observe that the latter equality implies that A is spectral if and only if there is a nondeterministic Turing machine that decides A in polynomial time in n—equivalently, if and only if there is a nondeterministic Turing machine that decides A in polynomial time in the size of its input, if the input is given in unary, that is, $n \in \mathbb{N}$ is represented as 1^n.

Durand et al. in [7] note that the set of spectral reals is closed neither under addition nor under multiplication (see [7, Corollary 3.11]) by proving the stronger result that there is a spectral real x such that none among $3x$, $x+1/3$, and x^2 is spectral (see [7, Theorem 3.10]). Related problems are well known for the larger class of computable reals where using the canonical base b-expansion as a representation of real numbers is generally eschewed because the basic algebraic operations fail to be uniformly computable (indeed, the operation $x \mapsto 3x$ is not computable if computable reals are represented by their canonical base-2 expansion)—even though the class of computable reals is a real closed field (see [23]). These problems were already identified by Turing [22] in 1937 and seem to have been anticipated as early as 1921 by Brouwer [4]. Instead of employing the canonical base-b expansion, a computable real number is usually specified in one of several equivalent ways, for instance, as computable sequences of nested intervals or as a computable Cauchy sequence of rational numbers with a computable modulus of convergence (see Weihrauch [23], Ko [18]). Using such representations it is possible to define proper complexity hierarchies of computable real numbers and computable functions on the reals, and it may be possible to define an alternative class of spectral reals using such representations by appealing to the Jones–Selman characterization above. We will not treat this possibility further in the paper.

1.1 Preliminaries In the remainder of the article we assume basic familiarity with computational complexity theory at the level of standard undergraduate textbooks such as Papadimitriou [20], Sipser [21], and Jones [14]. We set $\mathbb{N} = \{1, 2, 3, \ldots\}$, and for each finite alphabet we fix a standard recursive enumeration $M_1, M_2, \ldots$ of the set of multitape Turing machines with that alphabet. If M is the ith Turing machine in this enumeration, we define $\langle M \rangle = i$. Any Turing machine M defines a partial function $\phi_M : \{0,1\}^* \rightharpoonup \{0,1\}^*$. If this function is total, we write

$\phi_M : \{0,1\}^* \longrightarrow \{0,1\}^*$. Let $\mathbb{N}_{\{0,1\}^*} \subseteq \{0,1\}^*$ be the set of binary strings that represent elements of $\mathbb{N}$ in the usual way. If ϕ_M is total and $\phi_M(\mathbb{N}_{\{0,1\}^*}) \subseteq \mathbb{N}_{\{0,1\}^*}$, we write $\phi_M : \mathbb{N} \longrightarrow \mathbb{N}$; similarly, if ϕ_M is total and $\phi_M(\mathbb{N}_{\{0,1\}^*}) \subseteq \{0,1\}$.

Using the Jones–Selman characterization, it is evident that if a function $f : \mathbb{N} \longrightarrow \{0,1\}$ satisfies $f = \phi_M$ for some Turing machine M such that M computes $f(n)$ in time $2^{O(|\mathbf{b}(n)|)}$ for all n, then f is the characteristic function of the spectrum of a first-order sentence, and thus any real of the form $a + 0.f(1)f(2)f(3)\cdots$ with $a \in \mathbb{Z}$ is spectral. Observe that if f is computable in time polynomial in $|\mathbf{b}(n)|$, then a fortiori the real number $a + 0.f(1)f(2)\cdots$ is spectral.

2 Answers to the Open Problems

With the Jones–Selman characterization, the answers to all the open problems above are furnished simply by constructing sufficiently fast algorithms for computing the spectra in question.

2.1 Algebraic reals are spectral Recall that an algebraic number is a root of a nonzero polynomial with integer coefficients (see Hardy and Wright [11]). An *algebraic real* is a real algebraic number. The following is easily proved using the Jones–Selman characterization.

Proposition 2.1 *Every algebraic real is spectral.*

Proof Let α be an algebraic real, and write $\alpha = a + 0.b_1b_2\cdots$, where $a \in \mathbb{Z}$ and $0.b_1b_2\cdots$ is the canonical binary expansion of the fractional part of a. By standard results (see Hartmanis and Stearns [12, Theorem 11]), the function $f : \mathbb{N} \longrightarrow \{0,1\}$ with $f(n) = b_n$ can be computed in time polynomial in n using binary search, and hence α is spectral. □

2.2 Automatic reals are spectral We recall the following notions from Allouche and Shallit [2]. Let $k, b \geq 2$ be integers. A sequence $(a_i) = a_1a_2\cdots \in \{0,\ldots,b-1\}^\omega$ is said to be (k,b)-*automatic* if there is a deterministic finite automaton with output (DFAO) M such that if $x \in \{0,\ldots,k-1\}^*$ is the representation of the positive integer n in base k, then $M(x) = a_n$. A real number α is said to be (k,b)-*automatic* if there is $a_0 \in \mathbb{Z}$ and a (k,b)-automatic sequence (a_i) such that $\alpha = a_0 + \sum a_i b^{-i}$.

In their survey [7], Durand et al. more narrowly define automatic reals to be $(2,2)$-automatic in the above sense, and ask whether all such reals are spectral. Using the Jones–Selman characterization, it is immediate that it must be the case. If a real number $\alpha = a_0 + \sum a_i 2^{-i}$ is $(2,2)$-automatic, then (a_i) is computed by a suitable DFAO with input and output alphabet $\{0,1\}$. Any such DFAO can trivially be simulated by a deterministic multitape Turing machine using constant work space, hence in time polynomial in $|\mathbf{b}(n)|$ and thus a fortiori in time $2^{O(|\mathbf{b}(n)|)}$, whence (a_i) is the characteristic sequence of a spectrum.

We will prove a stronger and slightly more difficult result.

Proposition 2.2 *Let $k, b \geq 2$ be integers. Then, every (k,b)-automatic number is spectral.*

Proof Let $k, b \geq 2$, and let $\alpha = a + 0.b_1b_2\cdots$ with $a \in \mathbb{Z}$ be (k,b)-automatic. If α is rational, it is spectral, so in the remainder of the proof, we may assume that α is irrational. Observe that the expansion $\alpha = a + 0.b_1b_2\cdots$ is unique in this case.

As α is (k, b)-automatic, a result by Adamczewski and Cassaigne [1, Theorem 2.1] shows that α is not a Liouville number. Hence, there exists a positive integer m such that for all integers p, q with $q > 0$, we have $|\alpha - p/q| \geq q^{-m}$. We make use of this fact below.

Below, we describe a deterministic Turing machine M that, for all positive integers n, computes the prefix $b_1 \cdots b_n$ in time polynomial in n. It then follows that α is spectral. The construction of M is quite straightforward, hence we only give the high-level details.

Construction of M. On input 1^n, convert to binary $\mathbf{b}(n)$ and proceed as follows.

- $\mathbf{b}(1)$. M sets $c = 0$.
- $\mathbf{b}(n)$ for $n > 1$. M calls itself on input $\mathbf{b}(n-1)$ to obtain $b_1 \cdots b_{n-1}$, which yields the number $\sum_{i=1}^{n} b_i 2^{-i} = d/2^{n-1}$ for some integer d, and sets $c = d$.

  Observe that $b_n = 1$ iff $\alpha \geq (2c+1)/2^n$, and that M needs thus only check this inequality to output b_n.

  To check the inequality, M brute-force computes digits of the base-b expansions of α and of $(2c+1)/2^n$. Note that as α is not a Liouville number, we have $|\alpha - (2c+1)/2^n| \geq 2^{-nm} \geq b^{-nm}$.

  Hence, to ascertain whether $\alpha \geq (2c+1)/2^n$, at most the prefix of nm digits in the base-b expansions of α and $(2c+1)/2^n$ need to be computed by M.

  As α is automatic, the ith digit of the base-b expansion of α can be computed in constant space, hence in polynomial time in $|\mathbf{b}(i)|$, and as $(2c+1)/2^n$ is rational, the ith digit can clearly also be computed in polynomial time in $|\mathbf{b}(i)|$ and n. Thus, the total time needed to establish whether $\alpha \geq (2c+1)/2^n$ is polynomial in nm, hence polynomial in n.

End of construction of M. □

2.3 The subword complexity and entropy of spectral reals For a real number r, define $L_r \subseteq \{0,1\}^*$ to be the set of finite, nonempty bit strings occurring in the binary expansion of the fractional part of r (we choose the greedy binary expansion in case r is a dyadic rational). Furthermore, if $L \subseteq \{0,1\}^*$, set $L^n = L \cap \{0,1\}^n$ for all nonnegative integers n, and define $p_r(m) = |L_r^m|$.

In an earlier, unpublished version of [7] (see [8]), the authors defined the *binary string complexity* of a spectral number r to be the function $m \mapsto p_r(m)$. The survey authors asked ("Open Question 14") whether, for all spectral reals r, we would have $\lim_{m\to\infty} p_r(m)/2^m < 1$, or even $\lim_{m\to\infty} p_r(m)/2^m = 0$. This question can be easily answered in the negative by giving a (nondyadic) spectral number whose binary expansion is a *disjunctive sequence* in the sense of Jürgensen, Shyr, and Thierrin [16], [17], that is, it contains every binary string, and hence satisfies $\lim_{m\to\infty} p_r(m)/2^m = \lim_{m\to\infty} 2^m/2^m = 1$. The most well-known such number is Champernowne's binary constant (see [5]), obtained by concatenating the binary representations of the nonnegative integers together in sequence.

Proposition 2.3 *Champernowne's binary constant* $r = 0.11011100101110\cdots$ *is spectral. Hence, there is a spectral number* r *with* $\lim_{m\to\infty} p_r(m)/2^m = 1$.

Proof We will construct a deterministic Turing machine S that, for each $n \in \mathbb{N}$, on input 1^n computes the first n bits of r in time $O(n^2)$ (whence it is decidable in time

$O(n^2)$ whether the nth bit of r is 0 or 1). It then follows a fortiori from the Jones–Selman characterization that r is spectral. The construction of S is straightforward, whence we only give a high-level description.

Construction of S. Apart from its input and output tapes, S has two auxiliary tapes, each containing a counter. The first auxiliary tape counts the number of bits output so far, and S halts when the contents of this tape equal the value on the input tape. The second auxiliary tape contains another counter starting with the value 1. The Turing machine S writes the entire contents of this tape to the output, checking for each bit written to the output whether S should halt (by comparing the contents of the first auxiliary tape to the input tape). Thereafter, the counter on the second auxiliary tape is incremented by 1, and the process continues.

End of construction of S.

In the construction above, it is clear that for each bit *output*, S uses a number of steps proportional to its tape contents; as the number of cells used on both tapes is bounded above by the length of the original input, S uses at most $O(n^2)$ steps to output n bits. □

The limit $\lim_{m\to\infty} p_r(m)/2^m$ suffers from the deficiency that it cannot capture fine-grained variations in the growth rate of $p_r(m)$. In fact, for any real number r, if r is *not* disjunctive—that is, there is some binary word that does not occur in r—then $\lim_{m\to\infty} p_r(m)/2^m = 0$; this is easy to prove in the same way that one proves the well-known fact that the set of real numbers whose binary expansions do not contain some binary word w has Lebesgue measure zero. Thus, for *any* real number r, we have either $\lim_{m\to\infty} p_r(m)/2^m = 0$ or $\lim_{m\to\infty} p_r(m)/2^m = 1$, and both are possible for spectral numbers, as shown by the result for the Champernowne constant above and the fact that any rational number is spectral and satisfies $\lim_{m\to\infty} p_r(m)/2^m = 0$.

A more fine-grained notion of using subword counting to gauge the complexity of infinite strings is *subword entropy* (see Chomsky and Miller [6]), which is simply the entropy, H_r, of the function p_r; that is,

$$H_r = \lim_{m\to\infty} \frac{\log_2(p_r(m))}{m}$$

if the limit exists.

We will show momentarily that the limit $\lim_{m\to\infty} \frac{\log_2(p_r(m))}{m}$ always exists for any real number r. First we show a definition that will be of use several times in the remainder of the section.

Definition 2.1 (cf., e.g., [3]) A set $L \subseteq \{0,1\}^*$ is *factorial* if for every $u, v \in \{0,1\}^*$, $uv \in L$ implies $u, v \in L$. A set L is *right-extensible* if, for every $v \in L$, there is $w \in \{0,1\}^*$ such that $vw \in L$.

Observe that if L is both factorial and right-extensible, then every subword of $v \in L$ is also in L, and $v \in L$ implies that either $v0 \in L$ or $v1 \in L$.

Clearly, for any real number r, L_r as defined in Section 2.3 is both factorial and right-extensible as the elements of L_r are the words occurring in a right-infinite sequence. We then have the following.

Proposition 2.4 *Let r be a real number; then $H_r = \lim_{m\to\infty} \log_2(p_r(m))/m$ exists, and $H_r = \inf_m \log_2(p_r(m))/m$.*

Proof By standard results (see, e.g., Lind and Marcus [19, Lemma 4.1.7]), any sequence (a_n) of nonnegative reals such that $a_{m+n} \leq a_m + a_n$ satisfies that $\lim_m a_m/m$ exists and equals $\inf_m a_m/m$.

L_r is factorial, and hence $p_r(m+n) \leq p_r(m) \cdot p_r(n)$ for all nonnegative integers m and n, whence $\log p_r(m+n) \leq \log_2 p_r(m) + \log_2 p_r(n)$, and the result follows. $\square$

Clearly, $0 \leq \inf_m \frac{\log_2(p_r(m))}{m} \leq 1$ for any real r. For any rational number r, we have $H_r = 0$, and Champernowne's constant satisfies $H_r = 1$.

The main question of interest is then: *Which real numbers may occur as H_r for spectral r?* The remainder of this section is devoted to providing a partial answer to this question.

Proposition 2.5 *Let $r = a + \sum_n a_n 2^{-n}$ be a real number with $a \in \mathbb{Z}$ and $a_n \in \{0,1\}$ for all $n \in \mathbb{N}$. If there is a Turing machine M with $\phi_M(n) = a_n$ for all $n \in \mathbb{N}$, then there is a computable function $f : \mathbb{N}^2 \longrightarrow \mathbb{Q}$ such that $H_r = \inf_n \sup_k f(n,k)$.*

Proof For all nonnegative integers n, k, define $p_r^{\leq k}(n)$ to be the number of distinct subwords of length n among the first k bits in the binary expansion of r. Then, $p_r^{\leq k}(n)$ is increasing in k and $\lim_k p_r^{\leq k}(n) = p_r(n)$. Thus, $(\log p_r^{\leq k}(n))/n$ is increasing in k and $\lim_k(\log p_r^{\leq k}(n))/n = \log p_r(n)/n$. Define $f(n,k)$ to be the rational number $\sum_{i=1}^{k} a_i 2^{-i}$, where $a_1, \ldots, a_k$ are the initial k bits of $(\log p_r^{\leq k}(n))/n$. Clearly f is computable, and by Proposition 2.4 we have $H_r = \inf_n \sup_k f(n,k)$, as desired. $\square$

Clearly, any spectral r satisfies the requirement of Proposition 2.5.

The set of real numbers s such that $s = \inf_n \sup_k f(n,k)$ for a total computable function $f(n,k)$ is denoted by Π_2, and such reals s are called Π_2-*reals* (see Zheng and Weihrauch [24]). It is tantalizing to conjecture that every Π_2-real s can be realized as $s = H_r$ for a spectral r. We have been unable to prove this conjecture; instead, we have a weaker result, given in the following.

Recall from [24] that the set Π_1 of Π_1-*reals* (also called *right-computable reals*) consists of the real numbers s such that there exists a computable function $f : \mathbb{N} \longrightarrow \mathbb{Q}$ such that $s = \inf_n f(n)$. If we denote the ordinary set of computable real numbers in the sense of Turing [22] by Δ_1, it is known from [24] that $\Delta_1 \subsetneq \Pi_1 \subsetneq \Pi_2$.

The following lemma shows that every Π_1-real may be realized as the subword entropy of a spectral real.

Lemma 2.2 *If $s \in [0,1]$ is a Π_1-real, then there exists a spectral real r with $H_r = s$.*

Proof By a result of Hertling and Spandl [13, Theorem 22], for every Π_1-real $s \in [0,1]$ there exists a shift space with polynomial-time decidable language $L \subseteq \{0,1\}^*$ such that $s = \lim_m(\log |L^m|)/m$; it is well known that the language of any shift space is factorial and right-extensible (see Béal et al. [3]).

We will define a deterministic Turing machine M running in polynomial time that on input 1^n produces the first n bits of the expansion of a particular real number r; a fortiori it follows that r is spectral.

- M marks off n cells on a work tape ("Work Tape 1") and initially writes 0 in each of these cells. Denote the cells from left to right by $c_1 c_2 \cdots c_n$.
- For each index $i \in \{1, \ldots, n\}$ that is a square $i = k^2$ for some positive integer k, M considers the "block" consisting of the $(k+1)^2 - k^2$ cells $c_{k^2}, c_{k^2+1}, \ldots, c_{(k+1)^2-1}$ (if $(k+1)^2 - 1 > n$, the positions beyond n are ignored). Using a counter on another work tape ("Work Tape 2"), M then constructs the kth element, w, in the lexicographic order on $\{0,1\}^*$ on a separate work tape ("Work Tape 3"), and decides whether w is an element of L (observe that this takes time polynomial in n as L is decidable in polynomial time in $\mathbf{b}(n)$).

  If $w \in L$, the first $c_{k^2}, c_{k^2+1}, \ldots, c_{k^2+|w|-1}$ on Work Tape 1 are overwritten with w (observe that $|w| \leq (k+1)^2 - k^2 = 2k+1$ as the kth element of $\{0,1\}^*$ has length at most k). If $w \notin L$, nothing new is written on Work Tape 1.
- When all square indices $i = k^2$ have been processed, M outputs $c_1 \cdots c_n$.

Clearly, on input n, M always halts and outputs an element of $\{0,1\}^n$. For $j < n$, the output of M on input 1^j is clearly a proper prefix of the output of M on input 1^n. By construction, M considers at most n indices and for each index queries a decision procedure for L at most once. Hence, the total running time of M is polynomial in n, and hence M computes the characteristic sequence of a spectral real number r.

By construction, $L \subseteq L_r$, and hence $H_r \geq \lim_m (\log |L^m|)/m = s$. The remainder of the proof is devoted to showing that $H_r \leq s$.

Consider a "window" of m consecutive positions in the binary expansion of r, and let l be the index of the leftmost end of this window. Let w_l^m be the length-m word in this window; it is clear that $p_r(m) = |\bigcup_{l=1}^{\infty} \{w_l^m\}|$ and hence that $p_r(m) \leq |\bigcup_{l=1}^{m^2} \{w_l^m\}| + |\bigcup_{l=m^2+1}^{\infty} \{w_l^m\}|$.

Split on cases according to l as follows.

- For $l \leq m^2$, note that there are at most m^2 distinct words of length m in the initial prefix of length m^2 of the binary expansion of r. Hence, $|\bigcup_{l=1}^{m^2} \{w_l^m\}| \leq m^2$.
- For $l > m^2$, the construction of M entails that there are at least m consecutive zeros between each part of r potentially containing a word $v \in L^i$ with $i \in \mathbb{N}$. Any such word v must start from an index $j = k^2$ with $k \geq m$, and thus there are at least $k^2 - (k-1)^2 = 2k - 1 \geq k \geq m$ zeroes occurring immediately before v and $(k+1)^2 - k^2 = 2k+1 \geq i$ zeros occurring immediately after v. Hence, the "local part" of the binary expansion of r around such a v is of the form $0^m v 0^m$.

  Thus, we may write the binary expansion of r starting from the (m^2+1)th index as:

$$0^m 0^{j_1} v_1 0^m 0^{j_2} v_2 0^m \cdots .$$

  Consider a window of length m in the above and the length-m word, w, in this window. The set of all such words w can be partitioned into four types. We give the types below and for each type an upper bound on the number of distinct words of length m of each type that occur in the binary expansion of r.

Type I: $w = 0^m$. There is at most 1 word of Type I.

Type IIa: $w = 0^j v$, where $1 \leq j < m$ and v is a prefix of length $m - j$ of some word in L. As L is factorial, $v \in L^{m-j}$. Hence, for each j, there are at most $|L^{m-j}|$ words of Type IIa. As L is also extensible, we have $|L^{m-j}| \leq |L^m|$ for each j. Hence, there at most $m|L^m|$ words of Type IIa.

Type IIb: $w = v0^j$, where $1 \leq j < m$ and v is a postfix of length $m - j$ of some word in L. By reasoning exactly as in Type IIa, there are at most $m|L^m|$ words of Type IIb.

Type III: $w = 0^j v 0^k$, where $v \in L^i$ for some $1 \leq i \leq m$ and $j + k = m - i$. For each j, there are at most $(m - i)|L^j|$ words of Type III. As L is extensible, we have $|L^i| \leq |L^m|$, and there are hence at most $m^2|L^m|$ words of Type III.

Type IV: w is a proper subword of a word $v \in L$. In this case, as L is factorial and w has length m, we have $w \in L^m$. Hence, there are at most $|L^m|$ words of Type IV.

By the above, we have $|\bigcup_{l=m^2+1}^{\infty}\{w_l^m\}| \leq 1 + 2m|L^m| + m^2|L^m| + |L^m|$.

In summary, we have $p_r(m) \leq m^2 + 1 + 2m|L^m| + m^2|L^m| + |L^m|$ and hence $p_r(m) \leq 8m^2|L^m|$, whence

$$
\begin{aligned}
H_r &= \lim_m \bigl(\log p_r(m)\bigr)/m \leq \lim_m \bigl(3 + 2\log(m) + \log|L^m|\bigr)/m \\
&= \lim_m \bigl(\log|L^m|\bigr)/m = s,
\end{aligned}
$$

as desired. □

Noting that $\Delta_1 \subset \Pi_1$, we obtain by the above lemma the curious fact that all computable numbers in [0, 1] (thus, e.g., every automatic number in [0, 1], every algebraic number in [0, 1], and the fractional part of many of the standard constants such as π and e) occur as the subword entropy of a spectral real.

3 (Lack of) Closure Properties of the Spectral Reals

Durand et al. [7] have shown that the class of spectral reals is closed neither under addition nor under multiplication (see [7, Corollary 3.11]). The purpose of this section is to extend those results by showing that the set of spectral reals is not closed under operations from a very general class. In particular, Theorem 3.4 below affords a method for proving the existence of a spectral number x such that no $f_i(x)$ is spectral for certain infinite families $\{f_i : i \in I\}$ of functions.

Definition 3.1 Let $\mathbb{D}^{\geq 0}$ be the set of finite strings over $\{0, 1, .\}$ that contain "." exactly once, and let $\mathbb{D} = \mathbb{D}^{\geq 0} \cup (-\mathbb{D}^{\geq 0})$. An element $q \in \mathbb{D}$ is called a *representation of a dyadic rational* and corresponds to a dyadic rational $\lhd q \rhd$ in the straightforward way: $\lhd a_m \cdots a_0.b_1 \cdots b_n \rhd = \sum_{i=1}^{m} a_i 2^i + \sum_{j=1}^{n} b_j 2^{-j}$ and $\lhd - a_m \cdots a_0.b_1 \cdots b_n \rhd = -(\sum_{i=1}^{m} a_i 2^i + \sum_{j=1}^{n} b_j 2^{-j})$. The *length* $|d|$ of the representation d of a dyadic rational is just the length of the string d (in case of negative numbers, the sign $-$ is not counted toward the length).

Definition 3.2 A set of functions $\mathcal{F} = \{f_i : i \in I\}$, where I is either $\{1, \ldots n\}$ or $I = \mathbb{N}$ and with $f_i : [a, b] \longrightarrow \mathbb{R}$, is said to be a *spectral barrier on* $[a, b]$ if the following hold.

1. $\forall i \in I$: $f_i : [a, b] \longrightarrow \mathbb{R}$, and $f_i \in C^1([a, b])$.

2. $\forall i \in I$: $f_i' > 0$ on (a, b).
3. The function $G : I \times \mathbb{D} \times \mathbb{N}_0 \longrightarrow \mathbb{Z}$ that maps $(i, d, 0)$ to the integer part of $f_i(d)$ and maps (i, d, n) to the nth bit after the decimal separator of $f_i(d)$ is computable in time polynomial in $i + |d| + n$.
4. There is a deterministic Turing machine M that, on input 1^i and $y \in \mathbb{D}$ such that $\lhd y \rhd \in f_i([a, b])$, returns $t \in \{0,1\}^*$ with $yt \in \mathbb{D}$ and $\lhd yt \rhd \in f_i([a, b])$ such that the (by requirement 2 necessarily unique) $x \in [a, b]$ with $f(x) = \lhd yt \rhd$ is nondyadic.
   Furthermore, M runs in time polynomial in $i + |y|$.

In the above, requirement 4 is merely a formalization of the property that for every dyadic rational $\lhd y \rhd \in f_i([a, b])$, we may quickly compute another dyadic rational $\lhd yt \rhd$ in the image "close to" $\lhd y \rhd$ such that $f_i^{-1}(\lhd yt \rhd)$ is not dyadic.

Example 3.3 Let

$$\begin{aligned}\mathcal{F} &= \{x \mapsto kx : k \text{ integer, not a power of } 2\}\\ &\quad \cup \{x \mapsto x + 1/k : k \text{ integer, not a power of } 2\},\end{aligned}$$

and index $\mathcal{F}$ such that

$$f_i = \begin{cases} x \mapsto (ix)/2 & \text{for } i \text{ even and } i/2 \text{ not a power of 2,} \\ x \mapsto x + 2/(i+1) & \text{for } i \text{ odd and } (i+1)/2 \text{ not a power of 2,} \\ x \mapsto 3x & \text{for } i/2 \text{ or } (i+1)/2 \text{ a power of 2.} \end{cases}$$

Then, $\mathcal{F}$ is a spectral barrier on $[0, 1]$. It is clear that all the functions in $\mathcal{F}$ are differentiable with positive derivative on $(0, 1)$. To check requirement 3, note that the nth bit of $\lhd x \rhd + \lhd y \rhd$, $\lhd x \rhd \cdot \lhd y \rhd$, and $\frac{\lhd x \rhd}{\lhd y \rhd}$ for $\lhd y \rhd \neq 0$ can be computed using ordinary schoolbook addition, multiplication, and long division on $x, y \in \mathbb{D}$ by a *deterministic* Turing machine in time polynomial in $|x| + |y| + n$.

It is clear that the functions $x \mapsto x + 1/k$, where k is not a power of 2, are not going to cause trouble in requirement 4 in the definition of spectral barrier: all such functions send dyadic numbers to nondyadic numbers, so the preimage of a dyadic number will always be nondyadic. Thus it does not matter what M outputs on input $(\mathbf{b}(i), y)$ whenever i represents one of these functions.

Next we consider the functions $x \mapsto kx$, where k is not a power of 2. We write $k = b \cdot 2^j$, where b and j are integers with $b > 1$ odd. If $kx = \lhd y \rhd \in [0, 1]$ is a dyadic rational, we may write $kx = a/2^n$ with a and n integers and a odd. Then $x = a/b \cdot 2^{-n-j}$, and as both a and b are odd, x is a dyadic rational if and only if b divides a. If b does not divide a, we are done. If b divides a, then b does not divide $2a + 1$. Thus, for $\lhd y1 \rhd = a/2^n + 1/2^{n+1} = (2a + 1)/2^{n+1}$, the unique x' for which $kx' = \lhd y1 \rhd$ satisfies $x' = (2a + 1)/b \cdot 2^{-(n+1)-j}$ with $2a + 1$ and b both odd, whence x' cannot be rational. All the previous checks can certainly be performed in time polynomial in $|y|$ and i.

Theorem 3.4 *If $\mathcal{F}$ is a spectral barrier, then there is a spectral number r such that for every $f \in \mathcal{F}$, $f(r)$ is not spectral.*

Proof Let $\mathcal{F}$ be a spectral barrier on $[a, b]$. The first part of the proof is devoted to the case where $b > 0$.

Consider Turing machines N_i that have no output, but instead either halt in unique and distinct *accept* and *reject* states, or fail to halt. It is clear that such machines compute exactly the partial recursive functions $\{0,1\}^* \rightharpoonup \{0,1\}$, and that it is decidable in time polynomial in $|x|$ whether $x \in \{0,1\}^*$ is a correct representation of such a machine.

There is an algorithm that, for each $j \in \mathbb{N}$, on input 1^j constructs a Turing machine M_j as follows. If $\mathbf{b}(j)$ does not represent a Turing machine of the form as above, let M_j be a Turing machine that rejects its input immediately. If $\mathbf{b}(j)$ represents a Turing machine N_j of the correct form, let M_j be a Turing machine with the following behavior:

$$\phi_{M_j}(x) = \begin{cases} \phi_{N_j}(x) & \text{if } N_j \text{ halts after at most } 2^{2^{|x|\cdot j}} + 1 \text{ steps,} \\ 0 & \text{otherwise.} \end{cases}$$

The function $j \mapsto \langle M_j \rangle$ is clearly computable in polynomial time. The standard map taking 1^j to $\mathbf{b}(j)$ is computable in polynomial time, the check for $\mathbf{b}(j)$ having the correct form can be performed in linear time, and the machines M_j can be constructed in polynomial time by noting that new Turing machine states for computing the "clocking function" $|x| \mapsto 2^{2^{|x|\cdot j}} + 1$ can be added to N_j in time $O(j)$.

We now use the above construction to devise a deterministic multitape Turing machine T running in polynomial time, with $\phi_T : \mathbb{N} \longrightarrow \mathbb{D}$, and we set $r = \lim_n \phi_T(n)$.

The construction of T is fairly involved due to the need for successive approximations to real numbers and case splits for each approximation. We give full details below.

Construction of T. T has an input tape, an output tape, and five work tapes that we name as follows: a counter tape (used to contain a pair of natural numbers), an M_j-tape, a simulation tape, an r-tape (used to contain a representation of a finite prefix of r), and a computation tape.

Choose $a', b' \in \mathbb{D}$ such that (i) $\max(0, a) < \lhd a' \rhd < \lhd b' \rhd < b$, (ii) $|a'| = |b'|$, and (iii) a' and b' differ only in the rightmost (i.e., the least significant) bit.

On input 1^n with $n \in \mathbb{N}$, T does as follows.

- If $n = 0$, T returns a dyadic representation of $\lfloor \lhd a' \rhd \rfloor$.
- If $n > 0$, T halts when it has computed n bits after the decimal separator, and it returns this nth bit. We now describe T's computation on n in detail.
  1. T starts by writing a' on the r-tape and writing $(1, 1)$ on the counter tape. T never erases from the r-tape, so clearly $r \in [a', b']$. Define the total order on $<$ on $I \times \mathbb{N}$ by $(i_1, j_1) < (i_2, j_2)$ if $i_1 + j_1 < i_2 + j_2$ or if $i_1 + j_1 = i_2 + j_2$ and $i_1 < i_2$.
  2. T now uses the counter tape to count through $I \times \mathbb{N}$. When the counter tape is set to a value $(i, j) \in I \times \mathbb{N}$, T computes $\langle M_j \rangle$ and writes this on the M_j-tape. The r-tape holds a representation $r_{(i,j)}$ of a dyadic rational when the counter changed to (i, j); as no erasure occurs on the r-tape, we have $\lhd r_{(i,j)} \rhd \leq r$, that is, $\lhd r_{(i,j)} \rhd$ is a lower bound on r. Similarly, we obtain an upper bound as follows. Denote by $r^{(i,j)}$ the element of $\mathbb{D}$ obtained by adding one to the last bit of the number on the r-tape, and propagating carries leftward if necessary; clearly $\lhd a' \rhd \leq \lhd r_{(i,j)} \rhd < \lhd r^{(i,j)} \rhd \leq \lhd b' \rhd$. We let $y_{(i,j)}, y^{(i,j)} \in \mathbb{R}$ such that $y_{(i,j)} = f_i(\lhd r_{(i,j)} \rhd)$ and $y^{(i,j)} = f_i(\lhd r^{(i,j)} \rhd)$ (see Figure 1).

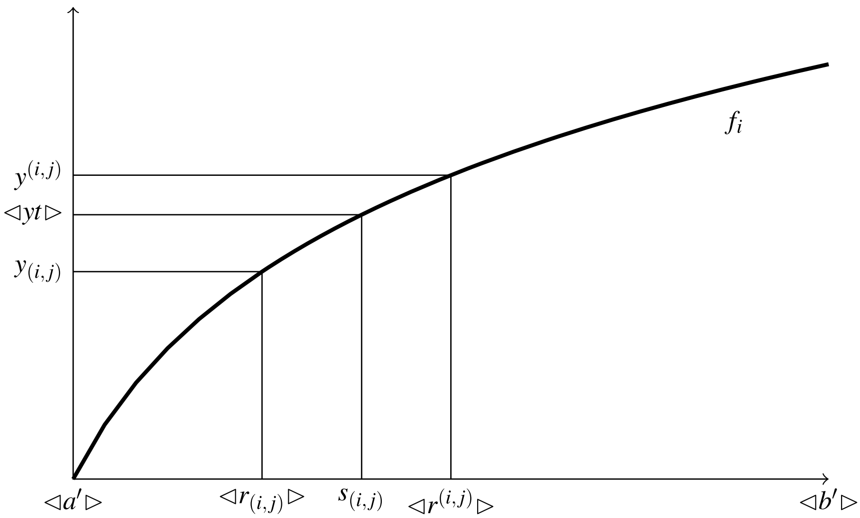

Figure 1 Rational approximations in the proof of Theorem 3.4.

3. T now finds a representation $y \in \mathbb{D}$ with $\lhd y \rhd \in (y_{(i,j)}, y^{(i,j)})$. To do this, T computes the integer parts of $y_{(i,j)}$ and $y^{(i,j)}$, and subsequently computes the first $2|r_{(i,j)}|$ bits of the canonical binary expansions of both $y_{(i,j)}$ and $y^{(i,j)}$. Denote by $y_{(i,j)}|_{2|r_{(i,j)}|}$ and $y^{(i,j)}|_{2|r_{(i,j)}|}$ the two elements of $\mathbb{D}$ obtained in this way, and observe that $y_{(i,j)}|_{2|r_{(i,j)}|}$ and $y^{(i,j)}|_{2|r_{(i,j)}|}$ can be found in polynomial time in $i + |r_{(i,j)}| + 2|r_{(i,j)}|$ by requirement 3 in the definition of spectral barrier.

   T then splits on cases as follows:

   "Success" If $|y^{(i,j)}|_{2|r_{(i,j)}|} - y_{(i,j)}|_{2|r_{(i,j)}|}| \geq 2 \cdot 2^{-2|\langle r_{(i,j)} \rangle|}$, T adds $3 \cdot 2^{-2|\langle r_{(i,j)} \rangle|-1}$ to $y_{(i,j)}|_{2|r_{(i,j)}|}$ to obtain $y \in \mathbb{D}$ such that for all strings $t \in \{0,1\}^*$, we have $\lhd yt \rhd \in (y_{(i,j)}, y^{(i,j)})$.

   "Failure" If $|y_{(i,j)}|_{2|r_{(i,j)}|} - y^{(i,j)}|_{2|r_{(i,j)}|}| < 2 \cdot 2^{-2|\langle r_{(i,j)} \rangle|}$, T writes a 0 on the r-tape. Then it updates $r_{(i,j)}$, $r^{(i,j)}$, $y_{(i,j)}$, and $y^{(i,j)}$ as if the counter just had increased to (i, j), and performs the above case split again.

   As $f_i \in C^1(a, b)$ and $f_i' > 0$, there is an $\varepsilon_i > 0$ such that $f_i' > \varepsilon_i$ on $[a', b']$. Then, $\lhd y^{(i,j)} \rhd - \lhd y_{(i,j)} \rhd \geq \varepsilon_i \cdot 2^{-|\langle r_{(i,j)} \rangle|}$, so eventually T will find some $y_{(i,j)}$ and $y^{(i,j)}$ such that the "Success" case above is encountered.

   Each case split above takes time polynomial in $|r_{(i,j)}|$, and after the "failure" case is encountered, it writes a bit on the r-tape. Hence, the entire process above uses at most time polynomial in n before the nth bit on the r-tape has been written.
4. Now T computes a string $t \in \{0,1\}^*$ such that the unique $s_{(i,j)} \in (\lhd r_{(i,j)} \rhd, \lhd r^{(i,j)} \rhd)$ with $f_i(s_{(i,j)}) = \lhd yt \rhd$ is nondyadic. This is

possible in time polynomial in $i + |y|$ by the definition of spectral barrier, hence possible in time polynomial in $|r_{(i,j)}|$.

5. Let $k_{(i,j)}$ be the position after the decimal separator of the rightmost 1 in yt. Now T computes $\langle M_j \rangle$ and writes it on the M_j-tape. T then simulates M_j on $k_{(i,j)}$ one step at a time. After each step in the simulation, T computes the next bit of $s_{(i,j)}$ and writes it on the r-tape.
To compute a bit of $s_{(i,j)}$, T simply uses binary search. First, let $r \in \mathbb{D}$ be the contents of the r-tape; T checks whether $f_i(\lhd r1 \rhd) \geq \lhd yt \rhd$. If it is, then the next bit of $s_{(i,j)}$ is 1; otherwise it is 0.

   The above check can clearly be performed in time polynomial in the number of bits of r computed, as $\lhd yt \rhd$ is dyadic and is not the image under f_i of a dyadic rational, and hence, in particular, is not equal to $f_i(\lhd r1 \rhd)$.

   If the simulation of M_j on $k_{(i,j)}$ accepts, T ensures that the $k_{(i,j)}$th bit of the dyadic representation of $f_i(\lhd r \rhd)$ is 0. That is, if $\lhd yt \rhd > 0$, T computes the bits of $s_{(i,j)}$ and writes them on the r-tape until the dyadic representation of $f_i(\lhd r \rhd)$ agrees with yt on all but the last nonzero bit of yt. Subsequently, T writes 0's on the r-tape until it has written 0 on a position where the bit in $s_{(i,j)}$ is 1. Similarly, if the simulation of M_j on $k_{i,j}$ rejects, T sets the $k_{(i,j)}$th bit of $f_i(r)$ to 1. Clearly, the total number of operations between each bit written on the r-tape is bounded above by some polynomial.

End of construction of T. From the assumptions about spectral barriers it is clear that T can be constructed such that the time it takes to compute the $(n+1)$th bit after the nth has been computed is only polynomial in n. Thus, T runs in time polynomial in the unary representation of the input, and thus r is spectral.

Assume for contradiction that the $f_i(r)$ were spectral for some $i \in I$. By the Jones–Selman characterization, there is a language $S \in \mathrm{NE}$ and an $a \in \mathbb{Z}$ such that $f_i(r) = a + \sum_{k \in S} 2^{-k}$. As $S \in \mathrm{NE}$, we can find a $c \in \mathbb{N}$ and an infinite number of deterministic Turing machines that decide S in time $2^{2^{nc}} + 1$; as there are infinitely many such machines N_j, we can choose one with $\langle N_j \rangle > c$. Now $\phi_{N_j} = \phi_{M_j}$ as N_j halts in time $2^{2^{nc}} + 1$. But by construction,

$$f_i(r) \neq a + \sum_{k \in L(M_j)} 2^{-k} = a + \sum_{k \in L(N_j)} 2^{-k} = a + \sum_{k \in S} 2^{-k} = f_i(r),$$

and we obtain the desired contradiction.

The above proves the case where $b > 0$. It remains to prove the theorem for the case $a < b \leq 0$. Here we observe that the set $\{x \mapsto -f_i(-x) : i \in I\}$ is a spectral barrier on $[-b, -a]$. By the proof for the case where $b > 0$, there is a real number $r \in (-b, -a)$ and a deterministic Turing machine T that on input 1^n computes the nth bit in r, and such that for all $i \in I$ no Turing machine running in time $2^{2^{nc}}$ computes the nth bit of $-f_i(-r)$ on input 1^n. As T is deterministic, we may compute the bits of a real number x as fast as the bits of $-x$, whence $-r \in (a, b)$ is spectral, but none of the $f_i(-r)$ are spectral. $\square$

Thus, by Example 3.3, there is a spectral real r such that the only elements $s \in \mathbb{N}r$ that are spectral are those for which $s = 2^k r$ with $k \in \mathbb{N}$.

References

[1] Adamczewski, B., and J. Cassaigne, "Diophantine properties of real numbers generated by finite automata," *Compositio Mathematica*, vol. 142 (2006), pp. 1351–72. Zbl 1134.11011. MR 2278750. DOI 10.1112/S0010437X06002247. 358, 368

[2] Allouche, J.-P., and J. Shallit, *Automatic Sequences: Theory, Applications, Generalizations*, Cambridge University Press, Cambridge, 2003. MR 1997038. DOI 10.1017/CBO9780511546563. 357

[3] Béal, M.-P., F. Mignosi, A. Restivo, and M. Sciortino, "Forbidden words in symbolic dynamics," *Advances in Applied Mathematics*, vol. 25 (2000), pp. 163–93. Zbl 0965.37014. MR 1780764. DOI 10.1006/aama.2000.0682. 359, 360

[4] Brouwer, L., "Besitzt jede reelle Zahl eine Dezimalbruchentwicklung?" *Mathematische Annalen*, vol. 83 (1921), pp. 201–10. MR 1512010. 356

[5] Champernowne, D. G., "The construction of decimals normal in the scale of ten," *Journal of the London Mathematical Society (2)*, vol. S1-8 (1933), pp. 254–60. MR 1573965. DOI 10.1112/jlms/s1-8.4.254. 358

[6] Chomsky, N., and G. A. Miller, "Finite state languages," *Information and Control*, vol. 1 (1958), pp. 91–112. Zbl 0081.14503. MR 0108417. 359

[7] Durand, A., N. D. Jones, J. Makowsky, and M. More, "Fifty years of the spectrum problem: Survey and new results," *Bulletin of Symbolic Logic*, vol. 18 (2012), pp. 505–53. Zbl pre06126491. MR 3053068. 355, 356, 357, 358, 362

[8] Durand, A., N. D. Jones, J. Makowsky, and M. More, "Fifty years of the spectrum problem: Survey and new results," preprint, arXiv:0907.5495v1 [math.LO]. 356, 358

[9] Ebbinghaus, H.-D., J. Flum, and W. Thomas, *Mathematical Logic*, 2nd ed., *Undergraduate Texts in Mathematics*, Springer, New York, 1994. MR 1278260. DOI 10.1007/978-1-4757-2355-7. 355

[10] Enderton, H. B., *A Mathematical Introduction to Logic*, 2nd ed., Harcourt/Academic Press, Burlington, Mass., 2001. MR 1801397. 355

[11] Hardy, G. H, and E. M. Wright, *An Introduction to the Theory of Numbers*, Oxford University Press, Oxford, 1938. 357

[12] Hartmanis, J., and R. E. Stearns, "On the computational complexity of algorithms," *Transactions of the American Mathematical Society*, vol. 117 (1965), pp. 285–306. Zbl 0131.15404. MR 0170805. 357

[13] Hertling, P., and C. Spandl, "Shifts with decidable language and non-computable entropy," *Discrete Mathematics and Theoretical Computer Science*, vol. 10 (2008), pp. 75–93. Zbl 1204.37013. MR 2457053. 360

[14] Jones, N. D., *Computability and Complexity from a Programming Perspective*, Foundations of Computing, MIT Press, 1997. Zbl 0817.68120. MR 1755599. 356

[15] Jones, N. D., and A. L. Selman, "Turing machines and the spectra of first-order formulas," *Journal of Symbolic Logic*, vol. 39 (1974), pp. 139–50. Zbl 0288.02021. MR 0485317. 356

[16] Jürgensen, H., H.-J. Shyr, and G. Thierrin, "Disjunctive ω-languages," *Elektronische Informationsverarbeitung und Kybernetik*, vol. 19 (1983), pp. 267–78. MR 0731304. 358

[17] Jürgensen, H., and G. Thierrin, "On ω-languages whose syntactic monoid is trivial," *International Journal of Computer and Information Sciences*, vol. 12 (1983), pp. 359–65. MR 0741785. DOI 10.1007/BF01008047. 358

[18] Ko, K.-I., *Complexity Theory of Real Functions*, Birkhäuser, Boston, 1991. MR 1137517. DOI 10.1007/978-1-4684-6802-1. 356

[19] Lind, D., and B. Marcus, *An Introduction to Symbolic Dynamics and Coding*, Cambridge University Press, Cambridge, 1995. Zbl 1106.37301. MR 1369092. DOI 10.1017/CBO9780511626302. 360

[20] Papadimitriou, C. H., *Computational Complexity*, Addison-Wesley, Reading, Mass., 1994. Zbl 0833.68049. MR 1251285. 356

[21] Sipser, M., *Introduction to the Theory of Computation*, PWS Publishing Company, 1996. 356

[22] Turing, A., "On computable numbers, with an application to the Entscheidungsproblem: A correction," *Proceedings of the London Mathematical Society*, vol. 43 (1937), pp. 544–46. MR 1575661. DOI 10.1112/plms/s2-43.6.544. 356, 360

[23] Weihrauch, K., *Computable Analysis: An Introduction*, Springer, Berlin, 1998. Zbl 0956.68056. MR 1795407. DOI 10.1007/978-3-642-56999-9. 356

[24] Zheng, X., and K. Weihrauch, "The arithmetical hierarchy of real numbers," *Mathematical Logic Quarterly*, vol. 47 (2001), pp. 51–65. Zbl 0968.03075. MR 1808946. DOI 10.1002/1521-3870(200101)47:1<51::AID-MALQ51>3.0.CO;2-W. 360

Acknowledgments

We thank Neil D. Jones and Søren Riis for useful comments, and Jeffrey Shallit for pointing us to the paper [1].

Jakobsen
Department of Mathematical Sciences
University of Copenhagen
Universitetsparken 5
DK-2100 Copenhagen Ø
Denmark
S.K.Jakobsen@qmul.ac.uk

Simonsen
Department of Computer Science
University of Copenhagen (DIKU)
Njalsgade 128–132
DK-2300 Copenhagen S
Denmark
simonsen@diku.dk